Hayya, J.C., Kim, J.G., Disney, S.M., Harrison, T.P. and Chatfield, D., (2006), "Estimation in supply chain inventory management", **International Journal of Production Research**, Vol. 44, No. 7, pp1313–1330. DOI: 10.1080/00207540500338039.

Estimation in supply chain inventory management

Jack C. Hayya*
Department of Supply Chain and Information Systems
The Smeal College of Business Administration
Pennsylvania State University
303 Beam Bldg
University Park, PA 16802, USA
Tel: 814-865-1461
E-mail: jch@psu.edu
Fax: 814-863-2381

Jeon G. Kim
Young Tong Maintenance and Operation Office
Korea Highway Corporation
Youngtong-Kun, Choong Chung Buk Do
Korea
Tel: 82-43-740-7300
E-mail: jxk125@hotmail.com
Fax: 82-43-740-7308

Stephen M. Disney
Logistics Systems Dynamics Group
Cardiff Business School
Cardiff University
Aberconway Building, Colum Drive
Cardiff, CF10 3EU, UK.
Tel: + 44 29 2087 6310
E-mail: disneysm@cardiff.ac.uk
Fax: +44 (0)29 2087 4301

Terry P. Harrison
Department of Supply Chain and Information Systems
The Smeal College of Business
Pennsylvania State University
Room 509-L BAB
University Park, PA 16802, U.S.A.
Tel: 814-863-3357
E-mail: tharrison@psu.edu
Fax: 814-683-7067

Dean Chatfield
Department of Management Science and Information Technology
Virginia Tech
1007 Pamplin Hall (0235)
Blacksburg, VA 24061, USA
Tel: 540-231-4593
E-mail: Deanc@vt.edu
Fax: 540-231-3752

*Corresponding author.

Estimation in supply chain inventory management

Abstract

Differences in estimation or forecasting procedures could produce dramatically different parameter estimates in supply chain inventory management. We show, for example, that determining when to introduce estimates of lead times in the calculation of the variance of demand during lead time can yield dramatically different safety stocks and order-up-to-levels. Also, calculations of supply chain variance amplification using a firewalled, sequential chain execution differ markedly from an analysis that considers a *k*-echelon analysis as a whole, $k > 2$. There is also the issue of forecasting lumpy demand when negative orders are not allowed. Our research compares the results in the recent literature and shows how apparently equivalent estimation procedures concerning demand during lead time (for example, using separate historical lead time and demand rate data versus directly using historical data of demand during lead time) are not equivalent; also, that the conventional exponential smoothing forecasting may not be appropriate at the higher echelons of supply chains where lumpy demand frequently occurs.

Keywords: supply management, forecasting, lead times, demand during lead time.

1. Introduction

As Zhang (2004), for example, has found, it makes a difference how we forecast supply chain demand in terms of detail. It makes a difference whether we introduce lead times explicitly into the analysis or whether we imply it in the estimation of the quantity called 'demand during lead time (*LTD*)'. It makes a difference if we used simple exponential smoothing (*SES*) for forecasting, rather than moving averages (*MA*) or some other method. It makes a difference in the computation of the bullwhip effect (*BWE*) if the lead time were greater or less than the length of the moving average. Also, it makes a difference if we insulated adjacent pairs of echelons by not allowing shortfalls or breakdowns to be transmitted upward in the supply chain. And the forecasting of lumpy demand, which is characteristic of the higher echelons of a supply chain when negative orders are not allowed, poses problems that may not be readily resolved by conventional forecasting, such as *SES* or *MA*. In the context of an *AR(1)* demand process, Zhang studied the impact of forecasting methods, such as moving averages and exponential smoothing, on the bullwhip effect and found that these forecasting methods affected average inventory costs. He also found that, for demand, a positive autocorrelation favored moving averages and a negative autocorrelation favored exponential smoothing. Our present paper focuses on *iid* demands, *i.e.*, an autocorrelation of zero.

The paper is organized as follows. Section 2 comments upon and reviews work on estimation procedures in the simulation and analysis of the *BWE* effect. Section 3 is a survey of several methods to forecasting lumpy demands, which, in our case, are random non-negative realizations separated by a random number of zeroes. In this section we use as an example a data set of 180 industrial observations from Croston (1972). Section 4 summarizes.

2. Review

Lead-time demand

Defining lead-time demand correctly is fundamental. *LTD* is the sum of the (variable) demands over the (variable) lead times. For example, suppose that period demand followed a gamma distribution with mean 4 and standard deviation 2, that is, period demand is $G\ (4, 2^2)$, and suppose that the lead time followed a Poisson distribution with mean 3, *i.e., P(3)*. Some typical realizations of *LTD* would be as shown in table 1.

Table 1. Lead-time demand is a random sum of random variables

Lead time in periods	Demand in 1st period	Demand in 2nd period	Demand in 3rd period	Demand in 4th period	Demand in 5th period	LTD
3	3.2	4.2	4.4	--	--	**11.8**
5	2	2	0.7	5.8	2.3	**12.8**
4	2.3	2.4	6.3	5.3	--	**15.3**

By simulation with unit demand gamma-distributed with mean 4 and standard deviation 2and with lead time Poisson-distributed with mean 3.

We see that *LTD* is a random sum of random variables, and the realizations 11.8, 12.8, 15.3 in table 1 are three values of what would constitute the *LTD* distribution if we continued to generate random lead times and their corresponding random period demands. Let *D* denote period demand and *L* the lead time; then the *LTD* may be defined as

$$X_t = \sum_{j=0}^{L-1} D_{t+j}\ , \tag{1}$$

but, for forecasting purposes, we need to define *LTD* as

$$X_t = \sum_{j=0}^{L-1} D_{t-j}\ , \tag{2}$$

which for statistical analysis is mathematically equivalent to (1).

Let $E(L) = \mu_L, Var(L) = \sigma_L^2, E(D) = \mu_D$, and $Var(D) = \sigma_D^2$. Now if *D* and *L* are independent, then

$$\begin{aligned} E(X) &= \mu_L \cdot \mu_D \\ &= L \cdot \mu_D, \text{ when } L \text{ is constant,} \end{aligned} \tag{3}$$

and

$$\begin{aligned} Var(X) &= \mu_L \sigma_D^2 + E(D)^2 \sigma_L^2 \\ &= L \cdot \sigma_D^2, \text{ when } L \text{ is constant.} \end{aligned} \tag{4}$$

But when both *L* and *D* are constant, we may write (3) as

$$X = L \cdot D, \tag{5}$$

which could mislead in that one may now consider *LTD* in the stochastic case to be the product of *L* and *D*, in which case, when *L* and *D* are independent, the mean *LTD* would be (3) and the variance becomes

$$Var(X) = \sigma_L^2\sigma_D^2 + \mu_L^2\sigma_D^2 + \mu_D^2\sigma_L^2, \tag{6}$$

which could be substantially larger than the correct *LTD* variance in (4). For example, let $L \sim P(3)$ and $D \sim G(4, 2^2)$, then $Var(X)$ under (4) is 60, whereas under (6), it is 96, and thus the safety stock for the same nominal service level would be erroneously inflated by a factor of $\frac{\sqrt{96}}{\sqrt{60}} = 1.26$.

Lead times

Where lead times are introduced in the analysis makes a difference. For example, Chen *et al.* (2000a) introduce lead times *explicitly* in the estimation of order-up-to levels, whereas Kim *et al.* (2005) instead do a *p*-period moving average *(MA(p))* of *LTD*, which *implicitly* accounts for the lead times. Thus, Chen *et al.* employ a constant lead time in calculating the order-up-to level,

$$S_t = L\overline{D}_t + z \cdot s_{et}^L, \tag{7}$$

where $\overline{D}$ is an *MA* of previous realizations of demand, s_{et}^L is the standard deviation of the forecast error over the lead time, with $e_t = D_t - \overline{D}_t$ the one-period-ahead forecast error, and $z \sim N(0,1)$ the standard normal variable. Note that the lead time, *L*, in (7) is *explicitly multiplicative*. See also Disney and Grubbström (2004: 3421). Now, let the order placed at time *t* at any echelon (Chen *et al.* 2000a: 437, and Kim *et al.*, 2005) be

$$Q_t = S_t - S_{t-1} + D_{t-1}. \tag{8}$$

Then following Ryan (1997: 19), (8) may be rewritten as (for convenience denote $s_{et}^{L} = s_t$)

$$Q_t = (1 + L/p)D_{t-1} - (L/p)D_{t-p-1} + z\sqrt{L}(s_t - s_{t-1}), \tag{9}$$

where

$$s_t^2 = \frac{\sum_{i=1}^{p}(D_{t-i} - \overline{D_t})^2}{p-1} = \frac{\sum_{i=1}^{p} D_{t-i}^2 - \frac{\left[\sum_{i=1}^{p} D_{t-i}\right]^2}{p}}{p-1} \tag{10}$$

is the sample variance of the most recent p observations of period demand. To compare the *explicitly multiplicative* and *implicitly additive* lead time, let z = 0 in (9). Then, when D is independent, that equation can be expressed as

$$Q_t = (1 + L/p)D_{t-1} - (L/p)D_{t-p-1}. \tag{11}$$

Hence,

$$\begin{aligned} Var(Q_t) &= (1 + L/p)^2 Var(D) + (L/p)^2 Var(D) \\ &= Var(D)\left[1 + \frac{2L}{p} + \frac{2L^2}{p^2}\right], \end{aligned} \tag{12}$$

since $Cov(D_{t-1}, D_{t-p-1}) = 0$.

But instead of (7), Kim *et al.* (2005) use

$$S_t = \overline{X}_t + z \cdot s_t(X), \tag{13}$$

where

$$\overline{X}_t = \frac{\sum_{i=1}^{p} X_{t-i}}{p} = \frac{\sum_{j=0}^{L-1}\sum_{i=1}^{p} D_{t-i+j}}{p} = \sum_{j=0}^{L-1}\overline{D}_{t+j}, \tag{14}$$

and

$$\begin{aligned} s_t(X) &= \sqrt{\sum_{t=1}^{p}(X_t - \overline{X}_t)^2} \\ &= \sqrt{\frac{1}{p-1}\left[\sum_{i=1}^{p}\sum_{j=0}^{L-1}\left(D_{t-i+j} - \overline{D}_{t+j}\right)^2 + \sum_{i=1}^{p}\sum_{j\neq m}^{L-1}\left(D_{t-i+j} - \overline{D}_{t+j}\right)\left(D_{t-i+m} - \overline{D}_{t+m}\right)\right]}. \end{aligned} \tag{15}$$

In (14), the lead time is *implicitly additive*, in which case, since if we assumed that lead time were deterministic and $z = 0$, (8) becomes using (13)

$$\begin{aligned} Q_t &= \overline{X}_t - \overline{X}_{t-1} + z[s_t(X) - s_{t-1}(X)] + D_{t-1}, \\ &= \overline{X}_t - \overline{X}_{t-1} + D_{t-1} \\ &= \frac{1}{p}\left[X_{t-1} - X_{t-p-1}\right] + D_{t-1} \\ &= \frac{1}{p}\left[D_{t-1} + D_t + \ldots + D_{t+L-2} - D_{t-p-1} - D_{t-p} - \ldots - D_{t-p+L-2}\right] + D_{t-1}, \end{aligned} \quad (16)$$

using (1), or

$$= \frac{1}{p}\left[D_{t-1} + D_{t-2} + \ldots + D_{t-L} - D_{t-p-1} - D_{t-p-2} - \ldots - D_{t-p-L}\right] + D_{t-1}, \quad (17)$$

using (2). Thus, from either Eq. (16) or (17) and since the D's are independent,

$$\begin{aligned} Var(Q_t) &= \frac{1}{p^2}\sum_{i=1}^{L} 2Var(D_i) + Var(D_{t-1}) + \frac{2}{p}Cov(D_{t-1,}D_{t-1}) \\ &= \frac{2LVar(D)}{p^2} + Var(D) + \frac{2}{p}Var(D), \end{aligned}$$

for $p \geq L$. Hence,

$$Var(Q_t) = \sigma_D^2\left[1 + \frac{2}{p} + \frac{2L}{p^2}\right]. \quad (18)$$

But with $p < L$, the covariances within the bracketed expression in (16) or (17) vanish, and we are left with

$$\begin{aligned} Var(Q) &= \frac{2Var(D)}{p} + Var(D) \\ &= \sigma_D^2\left[1 + \frac{2}{p}\right], \end{aligned} \quad (19)$$

which approximates (18) with $p >> L$. Hence, ostensibly similar methods [Kim *et al.* (2005) *vs.* Chen *et al*. (2000a) and Dejonckheere *et al.* (2004)] could produce different variance amplification (*VA*), as may be seen in tables 2 and 3.

For between-node sequence, one may choose that events, such as ordering and the like, occur simultaneously at each node. Or one may model a 'staggered start' to enable sequential events, meaning that the modeler delays the start of successive stages by a slight amount of time to enable sequential actions. For example, the customer's actions would occur at times 0.0, 1.0, 2.0, ...; the retailer's actions at 0.001, 1.001, 2.001, ...; the wholesaler's actions occur at 0.002, 1.002, 2.002; and so on. What this may do is to increase the lead time ever so slightly and thus inflate the *BWE* somewhat.

Sequence of events

This is a modeling issue that could influence the estimation process. One must consider within-node and between-node sequence of events. Within node, one may define one of two sequences:

1. At the beginning of each period, *t,* the inventory manager observes the inventory level and places an order, Q_t, to raise the inventory position to S_t. After the order is placed, customer demand, D_t, occurs. This follows the sequence of events in Chen *et al*. (2000a: 437, 2000b: 271), which came from Ryan (1997: 14). This sequence is consistent with (8).
2. 2. At the beginning of each period, *t,* the inventory manager observes the demand, D_t, for that period, then places an order, Q_t, to raise the inventory position to S_t. That is, we would replace D_{t-1} in (8) by D_t. Hence,

$$Q_t = S_t - S_{t-1} + D_t, \tag{20}$$

but that would not make any difference in the calculation of $Var(Q_t)$.

Table 2. Variance amplification (*VA*): a comparison with no information sharing – deterministic lead time

(No Safety Stock, $z = 0$)

	Retailer *k = 1*	**Wholesaler** *k = 2*	**Distributor** *k = 3*	**Factory** *k= 4*
Kim *et al.* (2005)	1.17	1.37	1.60	1.87
Chen *et al.'s* (2000a) *LB*	1.68	2.82	4.74	7.97
Dejonckheere et al.'s (2004)	1.67	2.99	5.72	11.43
Multiplier	1.44 1.43	2.06 2.18	2.96 3.58	4.26 6.11

Source: Kim et al. (2005) and Dejonckheere et al. (2004, Table 3, p. 739). L = lead time, deterministic = 4; $D \sim N(50,20^2)$ *; LB = lower bound; k = supply chain echelon; 'multiplier' refers to the values in rows 2 and 3divided by those in row 1—thus, for example, 1.44=1.68/1.17 .*

Table 3. Variance amplification (*VA*): a comparison with information sharing – deterministic lead time

(Safety stock, z = 2)

	Retailer *k = 1*	**Wholesaler** *k = 2*	**Distributor** *k = 3*	**Factory** *k= 4*
Kim *et al.* (2005)	1.04	1.06	1.11	1.18
Chen *et al.'s* (2000a) *LB*	1.68	2.64	3.88	5.41
Dejonckheere *et al.'s* (2004)	1.67	2.61	3.83	5.32
Multiplier	1.62 1.61	2.49 2.46	3.50 3.45	4.58 4.51

Source: Kim et al. (2005) and Dejonckheere et al. (2004, Table 3, p. 739). L = lead time, deterministic =

4; $D \sim N(50,20^2)$ *; LB = lower bound; k = supply chain echelon; 'multiplier' refers to the values in rows 2 and 3divided by those in row 1—thus , for example, 1.61=1.67/1.04 .*

Information sharing

Disney *et al.* (2004) caution that information sharing may not have an effect in improving supply chain performance, as hoped, and that simply passing on information to businesses in a supply chain could be harmful. Nevertheless, the consensus is on the side of information sharing. For example, Dejonckheere *et al*. (2004) find that information sharing reduces the variance amplification of order quantities in supply chains, and, in a mixed-integer programming model, Dominguez and Lashkari (2004: 2136) find that 'significant savings may indeed be achieved when integrating the supply chain by means of information'. Mitra and Chatterjee (2004) model a one-warehouse, two-retailer system and show that costs can be reduced in using actual demand information in setting the order-up-to level at the warehouse. Some other papers on "information enrichment" have tackled the issue of cost implicitly. For example, Mason-Jones and Towill (1997) speak of the speedier response and the dampening of demand magnification phenomena due to information sharing.

In a numerical study of a model with one supplier and several identical retailers, Cachon and Fisher (2000: 1032) examine information sharing, finding that costs were 2.2 % lower on average with full information, with a maximum difference of 12.1%. These percentages appear to be modest, because we can show that inventory costs are a function of the variance of lead-time demand, and since information sharing reduces that variance, information sharing should have a significant impact on the overall cost of a supply chain. [Information can be defined as the lower bound of the reciprocal of the variance (Kendall and Buckland 1967: 138); alternatively information could be equated to 'precision', also the reciprocal of the variance (Cochran 1977: 103)].

Kim *et al.* (2004) posit that the increase in inventory cost due to system variance, lead time's in this case, is less than

$$\frac{D(b+h)}{2Q}\sigma^2, \tag{21}$$

where b and h are the unit shortage and holding costs. According to (21), if the variance increased by 50 percent, with other parameters remaining the same, so would the system cost. He *et al*. (2005) also find that the system costs are nearly linear in the standard deviation of lead time. In these scenarios, the demand rate was constant and, thus, the variability of lead time can be used as a surrogate for system variability. So, according to the results in table 4, the cost at the factory due to the absence of information sharing would be 179.0/52.5 = 3.41 times that when information is shared. The difference in the results between table 4 and tables 2 and 3 is that the latter two were based on deterministic lead time, whereas table 4 is based on stochastic lead time.

Table 4. Comparison of information sharing vs. no information sharing: standard deviation and variance amplification (*VA*)

	Std. dev.: information Sharing	*VA*: info sharing	Std. dev.: no information sharing	*VA*: no info sharing
Customer	20	1.00	20	1.00
Retailer	30.0	2.25	30.0	2.25
Wholesaler	38.7	3.74	51.1	6.52
Distributor	46.1	5.31	92.8	21.53
Factory	52.5	6.89	179.0	80.10

Source: Kim et al. (2005). D ~ N(50, 400); L gamma with mean 4 and variance 4

Sequential echelon pairing

Echelon pairing is another modeling issue that influences the parameter estimates, because some estimation methods of a continuous supply chain multi-stage structure make the assumption that each node pairing (customer-retailer, retailer-wholesaler, wholesaler-distributor, and so on) operates like it is decoupled or "firewalled" from the rest of the supply chain. [This is an example of what Venkateswaran and Son (2004) have called a ‘modelling approximation’]. We call the node pairings ‘sequential pairing’. But similar to an electrical grid, it does make a difference if we do or do not allow perturbations in a lower echelon to be transferred to a higher echelon. Such perturbation, for example, could be inventory shortages. In the *SISCO* simulations (Chatfield, 2001), such perturbation was allowed, but not in Chen *et al.* (2000a). So as seen in table 5, with no information sharing, Chatfield *et al.*’s (2004) results showed greater amplification at higher supply chain echelons than the analytical work of Chen *et al.* (2000a). Also, the bounds provided by Chen *et al*. (2000a) may have overlooked inter-node interactions in a supply chain. Chatfield *et al.* performed a ‘sequential-pairs’ simulation to test this interaction hypothesis. In that experiment, they broke the supply chain into four pairs (customer-retailer, retailer-wholesaler, wholesaler-distributor, distributor-factory), first simulating customer-retailer, observing the variance amplification, and using the results as the retailer’s ordering policy for retailer-wholesaler; and so on. The paired simulations were executed hierarchically, and we borrow the results from Chatfield *et al.* (2004) and show them in table 5, noting that sequential pairing gives a smaller *BWE* than a *k*-node scenario, $k > 2$, because the sequential pairs avoid the cascading effect of stockouts from an upstream pair. The results with information sharing mirror those in Chen *et al.* (2000a), as may be seen in table 6. In comparing with the control engineering work of Dejonckheere *et al.* (2004), Chatfield *et al.* (2004). mimicked their design, and the results were nearly identical, as may be seen in tables 7 and 8.

Table 5. SISCO simulation vs. Chen *et al.* (2000a) with no information sharing: echelon standard deviation and amplification ratio (boldface, in parentheses)

	Customer	Retailer	Wholesaler	Distributor	Factory
Chen *et al.* (2000a)	20.00	27.49 (**1.89**)	37.78 (**3.57**)	51.92 (**6.74**)	71.36 (**12.73**)
SISCO Simulation	19.99	27.55 (**1.90**)	40.01 (**4.01**)	60.27 (**9.09**)	93.13 (**21.70**)
“Sequential pairs” SISCO Simulation	20.01	27.63 (**1.90**)	37.91 (**3.59**)	51.82 (**6.70**)	71.70 (**12.84**)

Source: Chatfield et al. (2004). Demand rate ~ N(50, 20^2); protection time = L + R = 4 + 1 = 5, deterministic; MA(15)forecasting. The simulation was run for 20 replications of 5200 time periods (-200 for warm-up).

Table 6. SISCO simulation vs. Chen *et al.* (2000a) with information sharing: echelon standard deviation and amplification ratio (boldface, in parentheses)

	Customer	Retailer	Wholesaler	Distributor	Factory
Chen *et al.* (2000a)	20.00	27.24 (**1.89**)	35.90 (**3.22**)	44.72 (**5.00**)	53.75 (**7.22**)
"Sequential pairs" SISCO Simulation	19.99	27.54 (**1.90**)	36.07 (**3.26**)	45.05 (**5.08**)	54.24 (**7.36**)

Source: Chatfield et al. (2004). Demand rate ~ $N(50, 20^2)$; protection time = L + R = 4 + 1 = 5, deterministic; MA(15) forecasting. The simulation was run for 20 replications of 5200 time periods (-200 for warm-up).

Table 7. SISCO simulation vs. Dejonckheere *et al.* (2004) with no information sharing: echelon standard deviation and amplification ratio (boldface, in parentheses)

	Customer	Retailer	Wholesaler	Distributor	Factory
Dejonckheere *et al.* (2004)	10.0	12.90 (**1.665**)	17.30 (**2.993**)	23.91 (**5.718**)	33.81 (**11.43**)
***SISCO* Simulation**	10.02	12.93 (**1.67**)	17.33 (**2.99**)	23.96 (**5.72**)	33.88 (**11.43**)

Source: Chatfield et al.(2004). Demand rate ~ $N(100, 10^2)$; protection time = L + R = 4 + 1 = 5, deterministic; MA(15) forecasting. The simulation was run for 20 replications of 5200 time periods (-200 for warm-up).

Table 8. SISCO simulation vs. Dejonckheere *et al.* (2004) with information sharing: echelon standard deviation and amplification ratio (boldface, in parentheses)

	Customer	Retailer	Wholesaler	Distributor	Factory
Dejonckheere *et al.* (2004)	10.0	12.90 (**1.665**)	16.15 (**2.607**)	19.56 (**3.826**)	23.07 (**5.321**)
SISCO Simulation	10.02	12.93 (**1.67**)	16.19 (*2.61*)	19.62 (**3.83**)	23.15 (**5.34**)

Source: Chatfield et al. (2004). Demand rate ~ $N(100, 10^2)$; protection time = L + R = 4 + 1 = 5, deterministic; MA(19)forecasting. The simulation was run for 20 replications of 5200 time periods (-200 for warm-up).

Batching

Oftentimes, a supply chain echelon may want to take advantage of quantity discounts or of full truck load economies of scale. It may then happen that the actual quantity ordered could be a multiple of the order quantity calculated in Eq. (8). Also proven by Caplin (1985: 1403) that increasing order size increases the *BWE,* what this tactic does is to exacerbate this phenomenon (Potter and Disney 2005), and this batching contingency is seldom taken account of in supply chain estimation procedures. According to Potter and Disney, there are two types of batching: time-based and order-based. Time-based batching occurs when orders are received less frequently than they are placed; quantity-based batching has to do with the use of the Economic Order Quantity (*EOQ*) or from advantages that accrue from packing in bulk or from quantity discounts. Potter and Disney recommend a batch size that is a multiple of average demand, with the smallest batch size under the circumstances preferable. But if the EOQ is to be used in a stochastic inventory system, it should be the one with backlogging (Hax and Candea 1984: 138), because shortages would be inevitable.

3. Forecasting lumpy demand

When negative orders are not allowed, the demand at higher echelons of a supply chain is often lumpy in that a positive order, $D > 0$, is followed by a series of zero orders. We see these starting with wholesaler in table 9, generated by *SISCO* simulation (Chatfield 2001). See also Shultz (1987: 453) who says that 'sporadic demand frequently occurs at higher levels of multi-echelon inventory systems as a result of inventory-replenishment decisions at lower levels …'.

But instead of using simulation data, we give in table 10 an industrial example from Croston (1972: 297), and the Croston data comprise 180 observations of 29 positive demands and 151 zero demands. The behavior of that intermittent demand is as illustrated in figure 1A with the subset of the first 15 observations, with the spikes denoting positive demands and the flat lines at the bottom denoting zero demands. On the surface, simple exponential smoothing of these lumpy data, as seen in figure 1B, does not seem satisfactory, and this proves to be the case.

Croston suggested that we forecast separately the positive demands and the inter-arrival times. We shall do that here in the following counter-example to the notion that moving averages or exponential smoothing would be appropriate for forecasting supply chain policy parameters.

Table 9. *SISCO* simulation: lumpy demand at the higher echelons

Customer	Retailer	Wholesaler	Distributor	Factory
75.697	95.514	110.439	127.494	0.000
9.819	93.955	144.206	154.682	186.131
63.099	**0.000**	142.010	218.129	238.356
34.699	39.035	**0.000**	212.088	350.651
61.299	20.150	**0.000**	**0.000**	333.541
55.211	56.194	**0.000**	**0.000**	**0.000**
44.458	54.564	14.055	**0.000**	**0.000**
14.947	25.368	52.073	**0.000**	**0.000**
84.945	**0.000**	**0.000**	**0.000**	**0.000**
46.410	97.177	**0.000**	**0.000**	**0.000**
93.560	34.758	80.994	**0.000**	**0.000**
35.667	117.229	10.928	**0.000**	**0.000**
27.310	29.294	159.937	**0.000**	**0.000**
50.202	7.793	17.706	173.543	**0.000**
42.151	41.734	**0.000**	9.320	**0.000**
77.120	21.715	**0.000**	**0.000**	**0.000**
29.395	77.037	**0.000**	**0.000**	**0.000**
71.682	60.238	20.939	**0.000**	**0.000**
47.472	77.208	91.507	**0.000**	**0.000**
60.758	53.341	102.805	**0.000**	**0.000**
32.347	60.397	71.470	40.757	**0.000**
57.704	21.330	61.662	102.834	**0.000**
62.696	65.117	3.073	83.582	**0.000**
47.449	86.061	88.440	**0.000**	**0.000**
73.019	20.188	132.881	113.195	**0.000**
51.839	89.357	**0.000**	208.837	**0.000**
37.165	19.766	93.995	**0.000**	348.589
49.170	37.950	**0.000**	99.422	**0.000**

Negative orders not allowed; customer demand $N(50, 20^2)$*; lead time gamma (4,* 2^2*); MA(15).*
Source: Chatfield (2001).

Table 10. Croston's (1972: 297) example of intermittent demand
(sample size, n = 180; zeroes in periods of no positive demand)

Period	1	5	6	7	12	36	39	42	47	56	58	60	67	71	
Demand	5	3	2	5	3	6	6	5	2	1	2	1	3	2	
Time	72	73	78	84	89	94	98	104	110	130	133	155	159	170	176
Demand	4	6	2	3	2	6	3	4	2	6	3	6	2	4	5

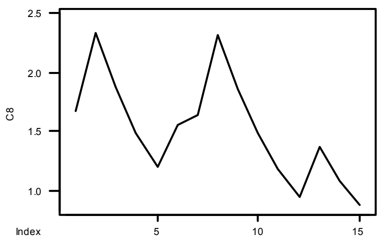

Figure 1A. Croston's data: actual Demand, first 15 Periods

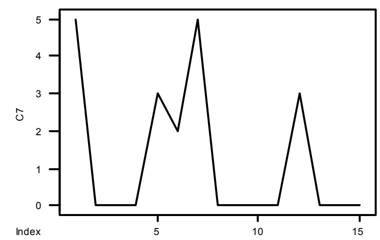

Figure 1B. Croston's data: simple exponential smoothing forecasts, first 15 periods
(smoothing constant, α = 0.2)

An experiment

We divide Croston's data into two sets of 90 periods each and use the first as the estimating or training set. The training set gives the estimates in table 11. We also determine that a smoothing constant α = 0.2 or 0.3 would give minimum mean absolute deviation (*MAD*) of the forecast errors, as seen in table 12. The importance of using *MAD* is that it is a surrogate for safety stocks, so that when the forecast errors are normally distributed, the standard deviation is (Silver *et al.* 1998: 112),

$$\sigma \approx 1.25 MAD. \tag{22}$$

MAD is an appropriate measure; it was the lone performance index in a simulation for forecasting lumpy demand by Bartezzaghi *et al.* (1999); and Venkateswaran and Son (2004) used it in examining the impact of modelling approximations in supply chains.

Table 11. Statistics from the training set
(First 90 observations of Croston's data)

Average inter-arrival interval	3.31 ≈ 3
Average size of positive demand (batch size)	3.32 units
Standard deviation of nonnegative demand	1.70 units
Time average of demands	0.70
Probability nonzero demand	0.21
Overall mean	0.70 units
Overall standard deviation	1.56 units
Coefficient of variation, *C*	2.23
Skewness	2.29 (vs. 3.10 theoretical).

Source: (Croston, 1972).

Table 12. Optimal smoothing constant
(first 90 observations of Croston's data)

α	***MAD***
0.023, obtained from fitting an ARIMA (0,1,1)	**1.35**
0.10	**1.16**
0.20	**1.13**
0.30	**1.13**
0.40	**1.14**

Naïve forecasting. When we forecast the second set of 90 observations, we use from table 11 an interval-arrival of 3 periods (*i.e.,* 3 zeroes) followed by a Demand of 3.32 (mean of the 1st 90 observations), as in this sequence:

0, 0, 0, 3.32, 0. 0, 0. 3.32, 0, 0, 0, 3.32, 0, 0, 0, 3.32, …

Astonishingly, the *MAD* for this naïve forecasting is **1.00 < 1.13** for the best α in Table 12.

All zeroes. Venkitachalam *et al.* (2002) suggest, among others, the approach of forecasting zeroes for all periods. This would give us a *MAD* = **0.52** and this method, as we shall see, would be appropriate when the probability of positive demand is small. The method carries to the limit Schultz's (1987: 454) suggestion of delaying the "placement of replenishment orders

… to achieve holding-cost savings that can outweigh the increased risk or cost of a stockout condition."

Random walk. In random walk forecasting (Makridakis *et al*. 1998: 329), today's realization is tomorrow's forecast. So when we use a random walk forecast, as in this sequence,

Period:	92	93	94	95	96	97	98	99	100	101	102	103	104	105
Demand:	0	0	0	6	0	0	0	3	0	0	0	0	0,	4 …,

we obtain *MAD* = **0.92.**

9-Period moving average. Since we shall be using a smoothing constant α = 0.2 for *SES*, we try an equivalent moving average, using (Brown, 1963: 108)

$$p = \frac{2}{\alpha} - 1 = 9.$$

The *MAD* here is **0.76**.

.**Static Monte-Carlo**. We use $P(D > 0) = 0.21$, with nonnegative $D \sim N(3.32, 1.72^2)$. This simulation yields *MAD* = **0.97.** We do this by first simulating a Bernoulli process with probability 0.21, which gives a string of 0's and 1's. Second, we randomize a normal variable, $N(3.32, 1.72^2)$, and substitute the simulated normal variates values for the 1's in the Bernoulli string.

Croston's algorithm, simplified. Croston advocates separating the lumpy data into two sets: one for positive demands, and another for the length of zero runs, which he calls 'interval'. Then he suggests using exponential smoothing on each set separately. We do this in table 13.

For the Croston algorithm, we used simple exponential smoothing on the positive demands and the intervals (or number of zero demands), with a smoothing constant, $\alpha = 0.2$, in both instances. When we put the forecasts against the actual realizations, we get *MAD* = **0.72.**

Table 13. A simplified Croston procedure
(second set of 90 observations from Croston’s data, smoothing constant, $\alpha = 0.2$)

D > 0	Interval forecast	Forecast for D > 0	Number zeroes
6	4	4.00000	7
3	4	4.40000	7
4	6	4.12000	6
2	6	4.09600	6
6	20	3.67680	6
3	3	4.14144	9
6	22	3.91315	8
2	4	4.33052	11
4	11	3.86442	9
5	6	3.89153	10

The Stuttering Poisson. The stuttering Poisson (*sP*) is a two-parameter distribution, with λt the average number of customers in a time interval *t*, and with each customer requesting an amount given by a geometric distribution with parameter, ρ (Ward 1978: 624). Now the average batch size is

$$1/(1-\rho), \tag{23}$$

and since the empirical batch size from table 11 is 3.32, this yields

$$\hat{\rho} = 0.70.$$

We get $\hat{\rho} = 0.55$ when, instead, we use the formula (Ward 1978: 625).

$$\rho = \frac{C^2 m - 1}{C^2 + 1}, \tag{24}$$

where

$$C = \sqrt{((1+\rho)/\lambda t)} \tag{25}$$

is the coefficient of variation of the *sP* over an interval *t*, and

$$m = \lambda t/(1-\rho) \tag{26}$$

is the mean. We estimate

$$\hat{\lambda} = \frac{2m}{(C^2 m + 1)} \text{ (Ward 1978: 625)}$$

$$= 0.31 \text{ (versus the empirical} \frac{1}{3.31} = 0.30 \text{ from table 11).}$$

The *sP* is made up of Poisson customer arrivals with parameter λ with each customer demanding a geometric order size of mean ρ. Now the geometric distribution is the discrete analog of the exponential. So we generate vectors of 90 Poisson arrivals and exponential order sizes, rounded to the nearest integer, and multiply these to arrive at random realizations of the *sP*, which we use as a forecast for the second set of 90 observations. Then for $\hat{\lambda} = 0.31$, $\hat{\rho} = 0.70$, we obtain

$$MAD = \mathbf{0.67},$$

and for $\hat{\lambda} = 0.31$, $\hat{\rho} = 0.55$, we obtain

$$MAD = \mathbf{0.62}.$$

A comparison

We compare in table 14 the various methods we used, employing *MAD* as the single metric. We find in our experiment that where all forecasts are zero gives the smallest *MAD*, followed by the stuttering Poisson (*sP*), Croston's algorithm, and *MA(9).* Astonishingly, simple exponential smoothing comes at the bottom of the list in table 14. Its *MSE* is 49 percent larger than the 'equivalent' *MA(9).* No wonder Dejonckheere *et al*. (2003: 581) found that the *BWE* 'generated by moving average forecasts in order-up-to model (therefore) in much less than that generated by exponential forecasts'. It seems counter-intuitive, because exponential smoothing is supposed to be 'generally superior' to moving averages (Makridakas *et al.* 1998: 145-146). But Sani and Kingsman (1997: 711, 712) also 'surprisingly' find in their study of the best periodic inventory control and demand forecasting for lumpy items that the moving average method outperforms other forecasting methods, particularly *SES*. And using thirteen forecasting methods to forecast lumpy-demand airline spare parts, Ghobbar and Friend (2003) argue that the choice of a forecasting method should depend on the degree of lumpiness. (We agree with their argument, as we elaborate in the next subsection.) But except for the Croston method and *SES*, the other forecasting methods in their menu included seasonality and linear trends, which we do not model in our experiment. Nevertheless, Ghobbar and Friend conclude that the use of *SES* for lumpy demand is questionable, a conclusion similar to that in Johnston and Boylan (1996), who demonstrated that *SES* was outperformed when the average inter-order arrival was greater than 1.25 forecast review periods, which could be interpreted to be $P(D > 0) \leq 0.44$. In our experiment that probability was 0.21.

But Chen *et al.* (2000b: 283-284) should cast away any doubts in proving that in supply chain forecasting *SES* is outperformed by an equivalent *MA(p)*. Let $Var(Q^{SES(\alpha)})$ and $Var(Q^{MA(p)})$ denote the respective variances with $\alpha = \frac{2}{p+1}$. Then, for *iid* demands with variance σ^2

$$\frac{Var(Q^{SES(\alpha)})}{\sigma^2} = 1 + \frac{4L}{p} + \frac{4L^2}{p(p+1)}, \tag{27}$$

greater than

$$\frac{Var(Q^{MA(p)})}{\sigma^2} = 1 + \frac{2L}{p} + \frac{2L^2}{p^2}. \tag{28}$$

Many have examined the conundrum of lumpy demand. For example, Venkitachalam *et al*. (2002) found that Croston's method coupled with bootstrapping (Davison and Hinkley 1999) yielded superior results. Prior to that, Willemain *et al*. (1994) found that Croston's method was 'robustly superior to exponential smoothing'. We add that it is not quite so simple: that the choice of method should depend on the degree of lumpiness as Ghobbar and Friend (2003) have asserted. In the present paper, we implied that that degree could be measured by *P(D > 0)*.`

Table 14. A comparison

Method	MAD	Comments
All Zeroes	**0.52**	All Forecasts put to zero
Stuttering Poisson	**0.67** **0.62**	Traditional characterization of lumpy demand
Croston's (simplified)	**0.72**	Use exponential smoothing separately on the set of positive demand and the set of intervals between.
MA(9)	**0.76**	Simple Moving Average that is equivalent to simple exponential smoothing with $\alpha = 0.2$
Random Walk	**0.92**	Today's realization is tomorrow's forecast
Static Monte Carlo	**0.97**	Use a Bernoulli string with the probability of positive demands
Naïve Forecasting	**1.00**	Using estimated from the1st set of 90 observations, repeat using an interval of 3 and a demand of 3.32
Simple Exponential Smoothing	**1.13**	Smoothing constant, $\alpha = 0.2$

MAD: Mean Absolute Deviation

The type of lumpy demand examined here consisted of zeroes/non-zeroes, with the non- zeroes stationary. We did not consider the type of lumpy demand characterized by sudden demand peaks as Miragliotta and Staudacher (2004) have done. And we are mindful as these authors have insisted that isolated techniques applied to forecasting procedures, rather than to inventory control, may be ineffective.

Degree of Lumpiness

Ghobar and Friend (2003) advance the common-sense hypothesis that, in forecasting lumpy demand, the method used should depend on the degree of lumpiness. We test this hypothesis in Chatfield and Hayya (2005), a factorial study where the degree of lumpiness is taken at three levels: high, with $P(D > 0) = 0.1$; mid, with $P(D > 0) = 0.5$; and low with $P(D > 0) = 0.9$. Included in the experiment at different levels are these factors: ordering cost, holding cost, shortage cost, and the coefficient of variation of positive demands. The criteria of goodness of forecast method were the following: a modified *MAPE* (mean absolute percentage error); *MSE* (mean square error); Theil's *U*-statistic (Makridakis *et al*., 1998:48); and inventory cost (ordering + holding + shortage). The forecast methods were the following: all-zero; the *MA*; *SES*; a Croston-type *MA*; the Croston *SES;* and the Stuttering Poisson (*sP*). See Table 14.

Indeed, we find Ghobar and Friend to be correct. With all factor levels operating, we find that all-zero forecasts outperform other methods in terms of forecasting error for high and mid-lumpiness when using *MAPE* as the criterion. In addition, the all-zero method produces the lowest cost when lumpiness is high, and also for mid-lumpiness if the unit shortage cost is greater than the unit holding cost, which is usually the case.

4. Summary

We should be aware that different procedures in estimation methods for supply chain inventory parameters could lead to dramatically different results, as we illustrate in this paper. For example if we considered lead times explicitly in the analysis, rather than subsuming it into the estimation of the mean and standard deviation of demand during lead time, then that would inflate the bullwhip effect and lead to excessive inventories. Also, the way we analyze a supply chain, for example, sequential pairs versus a holistic system, makes a difference, because sequential pairs build firewalls within the supply chain and these tend to attenuate any flash floods that could sweep through.

Also significant is the phenomenon of lumpy demand at the higher echelons of the supply chain. By that we mean that the demand series is characterized by one random string of positive demands, followed by a random string of zero demands, when negative demands (or reverse logistics) are not allowed. Exponential smoothing of the entire set of data performs poorly in this situation, and as the conventional forecasting method that sets the order-up-to levels is usually exponential smoothing, this could lead to inflation of the mean absolute deviation and of the safety stocks.

Acknowledgement

The authors acknowledge with thanks the research support from the Center for Supply Chain Research, Smeal College of Business, Pennsylvania State University.

References

BARTEZZAGHI, E., VERGANTI, R. and ZOTTERI, G., 1999. A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics,* 59(1-3), 499-510

BROWN, R.G., 1963. *Smoothing, forecasting and prediction* (Eglewood Cliffs, N.J.: Prentice-Hall).

CACHON, G.P. and FISHER, M., 2000. Supply Chain Inventory Management and the Value of Shared Information. *Management Science*, 46, 1032-1048.

CAPLIN, A.S., 1985. The variability of aggregate demand with (s, S) inventory policies. *Econometrica*, 53(6), 1395-1409.

CHATFIELD, D., 2001. SISCO and SCML- Software Tools for Supply Chain Simulation Modeling and Information Sharing. Unpublished Ph.D. Dissertation, Department of Management Science and Information Systems, Penn State University, University Park, PA 16803.

CHATFIELD, D., KIM, J. G., HARRISON, T. and HAYYA, J., 2004. The Bullwhip Effect - Impact of Stochastic Lead Times, Information Quality, and Information Sharing: a Simulation Study. *Production and Operations Management*, 13(4), 340-353.

CHATFIELD, D., HAYYA, J., 2005. All-Zero Forecasts for Lumpy Demand: a Factorial Study. Working Paper 05-2, Center for Supply Chain Research, Smeal college of Business, University Park, PA 16802.

CHEN, F., DREZNER, Z., RYAN, J.K. and SIMCHI-LEVI, D., 2000a. Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of forecasting, Lead times, and Information. *Management Science*, 46(3), 436-443.

CHEN, F., RYAN, J.K. and SIMCHI-LEVI, D., 2000b. The Impact of Exponential Smoothing Forecasts on the Bullwhip Effect. *Naval Research Logistics*, 47, 269-286.

COCHRAN, W.G. 1977. *Sampling Techniques* (3rd ed.) (New York: Wiley).

CROSTON, J. D., 1972. Forecasting and Stock Control for Intermittent Demand. *Operational Research Quarterly*, 23(3), 289-303.

DAVISON, A.C. and HINKLEY, D.V., 1999. *Bootstrap methods and their applications* (New York: Cambridge University Press).

DEJONCKHEERE, J., DISNEY, S.M., LAMBRECHT, M.R. and TOWILL, D.R., 2003. Measuring the Bullwhip Effect: A Control Theoretic Approach to Analyse Forecasting Induced Bullwhip in Order-Up-To Policies. *The European Journal of Operations Research*, 147(3), 567-590.

DEJONCKHEERE, J., DISNEY, S.M., LAMBRECHT, M.R. and TOWILL, D.R., 2004. The Impact of Information Enrichment on the Bullwhip Effect in Supply Chains: A Control Theoretic Approach. *European Journal of Operations Research*, 153(3), 727-750.

DISNEY, S.M. and GROBBSTRÖM, R.W., 2004. Economic consequences of production and inventory control policy. *International Journal of Production Research*, 42(17), 3419-3431.

DISNEY, S.M., NAIM, M.M. and POTTER, A., 2004. Assessing the impact of e-business on supply chain dynamics. *International Journal of Production Economics*, 89, 109-118.

DOMINGUEZ, H. and LASHKARI, R.S., 2004. Model for integrating the supply chain of an appliance company: a value of information approach. *International Journal of Production Research*, 42(11), 2113-2140.

GHOBBAR, A.A. and FRIEND, C.H., 2003. Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research*, 30(4), 2097-2114.

HAX, A.C. and CANDEA, D, 1984. *Production and Inventory Management* (Englewood Cliffs, NJ.: Prentice-Hall).

HE, X., KIM, J.G. and HAYYA, J., 2005. The cost of finite lead-time variability in inventory analysis. *International Journal of Production Economics*, 97(2), 130-142.

KENDALL, M.G. and BUCKLAND, W.R., 1967. *A dictionary of statistical terms* (New York: Hafner Publishing Company).

KIM, J.G., SUN, D., HE, X. and HAYYA, J., 2004. The (s, Q) Inventory Model with Erlang Lead Time and Deterministic Demand. *Naval Research Logistics*, 51 (6), 906-923.

KIM, J.G., CHATFIELD, D., HARRISON, T. and HAYYA, J., 2005. Quantifying the Bullwhip Effect in a Supply Chain with Stochastic Lead Time. *European Journal of Operational Research* (forthcoming).

JOHNSTON, F.R. and BOYLAN, J.E., 1996. Forecasting for items with intermittent demand. *Journal of the Operational Research Society*, 47, 113-121.

MAKRIDAKAS, S., WHEELWRIGHT, S.C. and HYNDMAN, R.J., 1998. *Forecasting methods and applications* (New York: Wiley).

MASON-JONES, R. and TOWILL, D.R., 1997. Information Enrichment: Designing the supply chain for competitive advantage. *Supply Chain Management*, 2(4), 137-149.

MIRAGLIOTTA, G. and STAUDACHER, A.P. , 2004. Exploiting information sharing, stock management and capacity oversizing in the management of lumpy demand. *International Journal of Production Research,* 42(13), 2533-2554.

MITRA, S. and CHATTERJEE, A.K., 2004. Leveraging information in multi-echelon inventory systems. *European Journal of Operational Research*, 152(1), 263-280.

POTTER, A. and DISNEY, S.M., 2005. Bullwhip and Batching: An exploration. *International Journal of Production Economics* (forthcoming).

RYAN, J.K., 1997. Analysis of inventory models with limited demand information. Ph.D. dissertation, Department of Industrial Engineering and Management Science, Northwestern University, Evanston, IL.

SANI, B. and KINGSMAN, B.G., 1997. Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48(7), 700-713.

SCHULTZ, C.R., 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society*, 38(5), 453-458.

SILVER, E. A., PYKE, D.F. and PETERSON, R., 1998. *Inventory management and Production Planning and Scheduling* (3rd ed.) (New York: John Wiley).

VENKITACHALAM, G.H.K., PRATT, D.B., YOUNG, C.F., MORRIS, S. and GOLDSTEIN, M.L., 2002. Forecasting and Inventory Planning for Parts with Intermittent Demand - A Case Study. http://anaconda.ecen.okstate.edu/publications/ERC2003-ADF-Final.doc

VENKATESWARAN, J. and SON, Y.J., 2004. Impact of modelling approximations in supply chain analysis – an experimental study. *International Journal of Production Research*, 42(15), 2971-2992.

WARD, J.B., 1978. Determining Reorder Points when Demand is Lumpy. *Management Science*, 24(6), 623-632.

WILLEMAIN, T.R., SMART, C.N., SHOKOR, J.H., and Desautels, P.A., 1994. Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10(4), 529-538.

ZHANG, X., 2004. The impact of forecasting methods on the bullwhip effect. *International Journal of Production Economics*, 88(1), 15-27.