

14 THE BULLWHIP EFFECT: MANAGERIAL INSIGHTS ON THE IMPACT OF FORECASTING AND INFORMATION ON VARIABILITY IN A SUPPLY CHAIN

**Frank Chen¹, Zvi Drezner², Jennifer K. Ryan³
and David Simchi-Levi⁴**

¹**Department of Decision Sciences,
National University of Singapore**

²**Department of Management Science/Information Systems,
California State University, Fullerton**

³**School of Industrial Engineering,
Purdue University**

⁴**Department of Industrial Engineering and Management Sciences,
Northwestern University**

14.1 INTRODUCTION

An important observation in supply chain management, popularly known as the **bullwhip effect**, suggests that demand variability increases as one moves up a supply chain. For example, empirical evidence suggests that the orders placed by a retailer tend to be much more variable than the customer demand seen by that retailer. This increase in variability propagates up the supply chain, distorting the pattern of orders received by distributors, manufacturers and suppliers.

As pointed out by Lee, Padmanabhan and Whang (1997 a, b), the term “bullwhip effect” was coined by executives of Procter and Gamble (P&G), the company which manufactures the Pampers brand of diapers. These executives observed that while the consumer demand for Pamper’s diapers was fairly constant over time, the orders for diapers placed by retailers to their wholesalers or distributors were quite variable, i.e., exhibited significant fluctuations over time. In addition, even larger variations in order quantities were observed in the orders that P&G received from its wholesalers. This increase in the variability of the orders seen by each stage in a supply chain was called the “bullwhip effect”. For a further discussion of this phenomenon, see, for example, Baganha and Cohen (1995), Caplin (1985), Kahn (1987), Kaminsky and Simchi-Levi (1996) and Sterman (1989).

The bullwhip effect is a major concern for many manufacturers, distributors and retailers because the increased variability in the order process (i) requires each facility to increase its safety stock in order to maintain a given service level, (ii) leads to increased costs due to overstocking throughout the system, and (iii) can lead to an inefficient use of resources, such as labor and transportation, due to the fact that it is not clear whether resources should be planned based on the average order received by the facility or based on the maximum order.

To better understand the impact of the bullwhip effect on an entire supply chain, consider the case of a simple, two stage supply chain consisting of a single retailer and a single manufacturer. The retailer observes customer demand and places orders to the manufacturer. To determine how much to order from the manufacturer, the retailer must **forecast** customer demand. Generally, the retailer will use the observed customer demand data and some standard forecasting technique to calculate these forecasts.

Next, consider the second stage in our supply chain, the manufacturer. The manufacturer observes the retailer's demand and places orders to his supplier. To determine these order quantities, the manufacturer must forecast the retailer's demand. In many supply chains, the manufacturer does not have access to the actual customer demand data. Therefore, he must use the orders placed by the retailer to perform his forecasting. If, as the bullwhip effect implies, the orders placed by the retailer are significantly more variable than the customer demand observed by the retailer, then the manufacturer's forecasting and inventory control problem will be much more difficult than the retailer's forecasting and inventory control problem. In addition, the increased variability will force the manufacturer to carry more safety stock or to maintain higher capacity than the retailer in order to meet the same service level as the retailer.

In this paper we discuss the causes of the bullwhip effect, as well as methods for reducing its impact. The analysis is based on the recent work by Chen, Drezner, Ryan and Simchi-Levi (1998), Chen, Ryan and Simchi-Levi (1998) and Lee, Padmanabhan

and Whang (1997 a, b). We start, in the following section, with a discussion of the five main causes of the bullwhip effect. We then present, in the second section, some new results which enable us to quantify the increase in variability due to two of these causes: demand forecasting and lead times. These results provide some useful managerial insights on controlling the bullwhip effect, which are discussed in the third section. We then extend these results, in the fourth section, to consider the impact of centralized demand information on the bullwhip effect. Finally, we conclude with a discussion of some methods for reducing the impact of the bullwhip effect.

14.2 IDENTIFYING THE CAUSES OF THE BULLWHIP EFFECT

It is clear that the bullwhip effect can lead to significant increases in costs and inventory levels throughout the supply chain. In order to control or eliminate the bullwhip effect we must first understand its causes. Lee, Padmanabhan and Whang (1997 a, b) identify five main causes of the bullwhip effect: demand forecasting, lead times, batch ordering, supply shortages and price variations. Caplin (1985) considers the impact of batch ordering on the bullwhip effect. Chen, Drezner, Ryan and Simchi-Levi (1998) and Chen, Ryan and Simchi-Levi (1998) consider the impact of demand forecasting and lead times on the bullwhip effect. In this section, we will briefly discuss how each of these factors can cause the bullwhip effect.

14.2.1 Demand Forecasting

Consider our example of a simple two stage supply chain in which each stage uses some form of demand forecasting to determine its desired inventory level and order quantity. Recall that a simple **order-up-to inventory policy** requires each stage of the supply chain to raise its inventory level up to a given target level in each period. One common form of this policy is to set the target inventory level in period t , y_t , equal to

$$y_t = \hat{\mu}_t^L + z\hat{\sigma}_t^L, \quad (14.1)$$

where $\hat{\mu}_t^L$ is an estimate of the mean lead time demand, $\hat{\sigma}_t^L$ is an estimate of the standard deviation of the forecast errors over the lead time, and the parameter z is chosen to meet a desired service level.

In order to determine its target inventory level, each stage of the supply chain must forecast both the expected demand and the standard deviation of demand. This forecasting can be done using any of a number of forecasting techniques, for example, moving average or exponential smoothing. In the moving average forecast, the forecast average demand per period is simply the average of the demands observed over some fixed number of periods, say p periods. In the exponential smoothing forecast, the forecast average demand per period is a weighted average of *all* of the previous demand observations, where the weight placed on each observation decreases with the age of the observation.

Any forecasting technique can cause the bullwhip effect. To understand this, note that one property of most standard forecasting methods is that the forecast is updated each time a new demand is observed. Therefore, at the end of each period, the retailer will observe the most recent demand, update her demand forecast based on

this demand, and then use this updated forecast to update her target inventory level. It is this updating of the forecast and order-up-to point in each period that results in increased variability in the orders placed by the retailer.

Finally, note that if the retailer follows a simple order-up-to inventory policy in which she does not update the desired inventory level in each period, then she would not see the bullwhip effect. In other words, if in every period the retailer places an order to raise the on-hand inventory to the same **fixed** level, then the orders seen by the manufacturer would be exactly equal to the customer demand seen by the retailer. Therefore, the variability in the orders seen by the manufacturer would be exactly equal to the variability of the customer demand, and there would be no bullwhip effect.

14.2.2 Lead Times

The lead time is defined as the time it takes an order placed by the retailer to be received at that retailer. Lead times can add to the bullwhip effect by magnifying the increase in variability due to the demand forecasting. To understand this, note that lead times increase the target inventory level, i.e., the longer the lead time, the larger the inventory level required. In addition, if as discussed above, the retailer updates her target inventory level in each period (using demand forecasting), then longer lead times will lead to larger changes in the target inventory level, and thus more variability in the orders placed by the retailer.

For example, if the demands seen by the retailer are independent and identically distributed (i.i.d.) from a normal distribution with mean μ and variance σ^2 , then an approximately optimal order-up-to inventory level in period t is

$$y_t = L\hat{\mu}_t + z\sqrt{L}\hat{\sigma}_t,$$

where L is the lead time plus 1, $\hat{\mu}_t$ is an estimate of μ , and $\hat{\sigma}_t$ is an estimate of σ . In this case, we clearly see that if, from time t to $t + 1$, our estimate of μ changes by Δ , then our order-up-to level will change by $L\Delta$, where $L \geq 1$. In other words, any changes in our estimates of the parameters of the demand process will be magnified by the lead time.

14.2.3 Batch Ordering

Another major cause of the bullwhip effect is the “batching” of orders. The impact of batch ordering is quite simple to understand: if the retailer uses batch ordering, then the manufacturer will observe a very large order, followed by several periods of no orders, followed by another large order, and so on. Thus the manufacturer sees a distorted and highly variable pattern of orders.

Caplin (1985) considers the impact of batch ordering on the bullwhip effect. He considers a retailer who follows a continuous review (s, S) inventory policy in which the retailer continuously monitors the inventory level, and when the inventory level drops to s , places an order to raise the inventory level to S . Note that for an inventory policy of this form, when the retailer places an order, the size of that order, $Q = S - s$, is fixed and known. In this case, since the size of the order is fixed, the variability of the orders placed by the retailer is due only to the variability in the time between

orders, i.e., the variability in the time it takes for the inventory level to fall from S to s . Caplin proves that, if the demands faced by the retailer are i.i.d., then the variance of the orders placed by the retailer is greater than the variance of the customer demand observed by the retailer, and that the variance of the orders increases linearly in the size of the orders, i.e., the variability will increase linearly in Q .

14.2.4 Supply Shortages

A fourth cause of the bullwhip effect is associated with anticipated supply shortages. If a retailer anticipates that a particular item will be in short supply, he may place an inflated, unusually large, order with his supplier, because experience suggests that the supplier will be rationing the product among his customers based on the size of their orders. This “rationing” or “gaming” distorts the true demand pattern and gives the manufacturer a false impression of the market demand for the product. Lee, Padmanabhan and Whang (1997 a, b) discuss in detail the impact of this type of gaming on the variability in a supply chain.

14.2.5 Price Variations

A final cause of the bullwhip effect is the frequent price variations seen throughout a supply chain. For example, many retailers will offer products at a regular retail price with periodic price promotions and clearance sales. Clearly, when the price for an item changes, the customer demand for that item will also change. For example, during a promotion the retailer will see higher than usual demand, while after the promotion, the retailer may observe unusually low demand. Therefore, periodic price promotions can cause distorted demand patterns and increased variability in demand.

This phenomenon is observed at other stages of the supply chain as well. For example, when a manufacturer offers a trade promotion, retailers may place unusually large orders and stockpile inventory, and may not order again for several periods. Again, this causes distorted demand patterns and increases the variability in demand.

14.3 QUANTIFYING THE BULLWHIP EFFECT

So far, we have discussed a number of the causes and effects of the bullwhip effect. In order to better understand and control the bullwhip effect, it would also be useful to **quantify** the bullwhip effect, i.e., quantify the increase in variability that occurs at every stage of the supply chain. This would be useful not only to demonstrate the magnitude of the increase in variability, but also to show the relationship between the demand process, the forecasting technique, the lead time and the increase in variability.

To quantify the increase in variability for a simple supply chain, consider a two stage supply chain with a retailer who observes customer demand in time t , D_t , and places an order, q_t , to the manufacturer. Assume the customer demands seen by the retailer are random variables of the form

$$D_t = \mu + \rho D_{t-1} + \epsilon_t, \quad (14.3)$$

where μ is a non-negative constant, ρ is a correlation parameter with $|\rho| < 1$, and the error terms, ϵ_t , are i.i.d. from a symmetric distribution with mean 0 and variance σ^2 .

Suppose that the retailer faces a fixed lead time, L , such that an order placed by the retailer at the end of period t is received at the start of period $t + L$. Also, suppose the retailer follows a simple order-up-to inventory policy in which, in every period, she places an order to bring her inventory level up to a target level. That target inventory level is as given in equation (14.1).

14.3.1 Moving Average Forecasts

In order to implement this inventory policy, the retailer must estimate the mean and standard deviation of demand based on her observed customer demand. Suppose the retailer uses one of the simplest forecasting techniques: the moving average. In other words, in each period the retailer estimates the mean leadtime demand as $\hat{\mu}_t^L = L\hat{\mu}_t$ where $\hat{\mu}_t$ is an average of the previous p observations of demand, i.e.,

$$\hat{\mu}_t = \frac{\sum_{i=1}^p D_{t-i}}{p}.$$

She estimates the standard deviation of the L period forecast errors using the sample standard deviation of the single period forecast errors, $e_t = D_t - \hat{\mu}_t$.

Our objective is to quantify the increase in variability. To do this, we must determine the variance of q_t relative to the variance of D_t , i.e., the variance of the orders placed by the retailer to the manufacturer relative to the variance of the demand faced by the retailer. For this purpose, we write q_t as

$$q_t = y_t - y_{t-1} + D_{t-1}.$$

Observe that q_t may be negative, in which case we assume, similarly to Kahn (1987) and Lee, Padmanabhan and Whang (1997b), that this *excess inventory is returned without cost*. This assumption is not a standard one in inventory models, but it is necessary to obtain analytical results. In addition, simulation results indicate that this assumption has little impact on the results presented in this paper. For a discussion of these simulation results, see Chen, Drezner, Ryan and Simchi-Levi (1998). For further discussion of this assumption see Lee, Padmanabhan and Whang (1997b)

Given the estimates of the mean and standard deviation of the lead time demand, we can write the order quantity q_t as

$$\begin{aligned} q_t &= \hat{\mu}_t^L - \hat{\mu}_{t-1}^L + z(\hat{\sigma}_t^L - \hat{\sigma}_{t-1}^L) + D_{t-1} \\ &= (1 + L/p)D_{t-1} - (L/p)D_{t-p-1} + z(\hat{\sigma}_t^L - \hat{\sigma}_{t-1}^L). \end{aligned}$$

And taking the variance of q_t we get

$$\begin{aligned} Var(q_t) &= (1 + L/p)^2 Var(D_{t-1}) + (L/p)^2 Var(D_{t-p-1}) \\ &\quad - 2(L/p)(1 + L/p)Cov(D_{t-1}, D_{t-p-1}) \\ &\quad + 2z(1 + 2L/p)Cov(D_{t-1}, \hat{\sigma}_t^L) \\ &\quad + z^2 Var(\hat{\sigma}_t^L - \hat{\sigma}_{t-1}^L) \\ &= \left[1 + \left(\frac{2L}{p} + \frac{2L^2}{p^2} \right) (1 - \rho^p) \right] Var(D) \end{aligned}$$

$$+2z(1+2L/p)Cov(D_{t-1}, \hat{\sigma}_t^L) \\ +z^2Var(\hat{\sigma}_t^L - \hat{\sigma}_{t-1}^L),$$

where the second equation follows from $Var(D) = \frac{\sigma^2}{1-\rho^2}$ and $Cov(D_{t-1}, D_{t-p-1}) = \frac{\rho^p}{(1-\rho^2)}\sigma^2$.

To further evaluate $Var(q_t)$ we need the following lemma:

Lemma 14.3.1 *If the retailer uses a simple moving average forecast with p demand observations, and if the demand process satisfies (14.3), then*

$$Cov(D_{t-i}, \hat{\sigma}_t^L) = 0 \text{ for all } i = 1, \dots, p.$$

We therefore have the following lower bound on the increase in variability from the retailer to the manufacturer:

Theorem 14.3.2 *If the retailer uses a simple moving average forecast with p demand observations, an order-up-to inventory policy defined in (14.1), and if the demand process satisfies (14.3), then the variance of the orders, q^{MA} , placed by the retailer to the manufacturer, satisfies*

$$\frac{Var(q^{MA})}{Var(D)} \geq 1 + \left(\frac{2L}{p} + \frac{2L^2}{p^2} \right) (1 - \rho^p), \quad (14.7)$$

where $Var(D)$ is the variance of customer demand. The bound is tight when $z = 0$.

Several observations can be made from this relationship. First, we notice that the increase in variability from the retailer to the manufacturer is a function of three parameters, (1) p , the number of observations used in the moving average, (2) L , the leadtime between the retailer and the manufacturer, and (3) ρ , the correlation parameter. The impact of the number of observations and the lead time is intuitive. The increase in variability is a decreasing function of p , the number of observations used in the moving average, and an increasing function of L , the lead time.

The correlation parameter, ρ , can have a significant impact on the increase in variability. First, if $\rho = 0$, i.e., if demands are independent and identically distributed (i.i.d.), then

$$\frac{Var(q^{MA})}{Var(D)} \geq 1 + \frac{2L}{p} + \frac{2L^2}{p^2}. \quad (14.8)$$

Second, if $\rho > 0$, i.e., if the demands are positively correlated, then the larger ρ , the smaller the increase in variability. In addition, we see that positively correlated demands lead to less variability than i.i.d. demands. On the other hand, if $\rho < 0$, i.e., if the demands are negatively correlated, then we see some strange behavior. For even values of p , $(1 - \rho^p) < 1$, while for odd values of p , $(1 - \rho^p) > 1$. Therefore, for $\rho < 0$, the lower bound on the increase in variability will be larger for odd values of p than for even values of p .

Finally, we note that the lower bound is tight in the case of $z = 0$, i.e., if the retailer's order-up-to level is just a multiple of the mean demand. We will discuss inventory

policies of this form later in this paper. In addition, simulation estimates indicate that this the lower bound is quite good for other values of z . In other words, the lower bound given in (14.7) accurately describes the behavior of the system even in cases where $z \neq 0$.

Figure 1 shows the lower bound (solid lines) on the increase in variability for $L = 1$ and $z = 2$ for various values of ρ and p . It also presents simulation estimates of the exact value of $Var(q_t)/\sigma^2$ (dotted lines). We see that the increase in variability is larger for small values of the correlation than for high values of the correlation. Figure 2 shows the lower bound (solid lines) on the variability amplification as a function of p for various values of the lead time, L , for $\rho = 0$ and $z = 2$. It also presents simulation estimates of the exact value of $Var(q_t)/\sigma^2$ (dotted lines). For additional simulation results, see Chen, Drezner, Ryan and Simchi-Levi (1998).

14.3.2 Exponential Smoothing Forecasts

We can perform a similar analysis for any forecasting method. Consider, for example, exponential smoothing, one of the most common forecasting techniques in practice. In this case, the retailer estimates the mean leadtime demand as $\hat{\mu}_t^L = L\hat{\mu}_t$ with

$$\hat{\mu}_t = \alpha D_{t-1} + (1 - \alpha)\hat{\mu}_{t-1},$$

where $0 < \alpha \leq 1$ is the smoothing constant. In other words, the current forecast, $\hat{\mu}_t$, is just the weighted average of the previous period's demand and the previous period's forecast demand, where α is the relative weight to be placed on the current period's demand. We note that, in general, $0 < \alpha \leq 0.30$. We assume that the retailer estimates the standard deviation of the forecast errors using the standard estimate based on the mean absolute deviation of the forecast errors. For more information on exponential smoothing forecasts, see Hax and Candea (1984, Chap. 4).

Given these estimates of the mean and standard deviation, the retailer can form an order-up-to inventory policy as defined in equation (14.1), and we can repeat the above analysis. In this case, if the demands seen by the retailer satisfy (14.3), we have the following lower bound on the increase in variability from the retailer to the manufacturer.

Theorem 14.3.3 *If the retailer uses a simple exponential smoothing forecast with smoothing parameter α , $0 < \alpha \leq 1$, and an order-up-to inventory policy defined in (14.1), then the variance of the orders, q^{EX} , placed by the retailer to the manufacturer, satisfies*

$$\frac{Var(q^{EX})}{Var(D)} \geq 1 + \left(2L\alpha + \frac{2L^2\alpha^2}{2 - \alpha} \right) \left(\frac{1 - \rho}{1 - \beta\rho} \right), \quad (14.10)$$

where $Var(D)$ is the variance of customer demand. The bound is tight when $z = 0$.

Again, several important observations can be made from this relationship. First, we notice that the increase in variability from the retailer to the manufacturer is a function of three parameters, (1) α , the smoothing parameter used in the exponential

smoothing, (2) L , the leadtime between the retailer and the manufacturer, and (3) ρ , the correlation parameter.

The impact of the smoothing parameter and the lead time is quite intuitive. The increase in variability is an increasing function of α , the smoothing parameter. This can be explained as follows. Recall that α is the weight placed on the most recent observation of demand in the exponential smoothing forecast. Therefore, the more weight we place on a single observation, the greater the increase in variability. Also notice that, as with the moving average, the increase in variability is an increasing function of L , the lead time.

The correlation parameter, ρ , can also have a significant impact on the increase in variability. First, if $\rho = 0$, i.e., if demands are independent and identically distributed, then

$$\frac{Var(q^{EX})}{Var(D)} \geq 1 + 2L\alpha + \frac{2L^2\alpha^2}{2 - \alpha}. \quad (14.12)$$

Next, notice that if $0 \leq \rho < 1$,

$$\frac{1 - \rho}{1 - \beta\rho} \leq 1,$$

while if $-1 < \rho \leq 0$,

$$\frac{1 - \rho}{1 - \beta\rho} \geq 1.$$

Therefore, for **positively correlated** demands, the increase in variability will be **less** than for i.i.d. demands ($\rho = 0$). On the other hand, for **negatively correlated** demands, the increase in variability will be **greater** than for i.i.d. demands.

Finally, we again note that the lower bound is tight in the case of $z = 0$, and that simulation estimates indicate that this the lower bound is quite good for other values of z . In other words, the lower bound given in (14.3.3) accurately describes the behavior of the system even in cases where $z \neq 0$.

Figure 3 presents the increase in variability for a retailer using an exponential smoothing forecast when $L = 1$ and $z = 0$ for various values of the correlation parameter, ρ . The figure demonstrates that negatively correlated demands can lead to significantly higher variability than positively correlated demands. For additional results see Chen, Ryan and Simchi-Levi (1998).

14.3.3 Multiple Retailers and a Single Manufacturer

The model presented above clearly does not capture many of the complexities involved in real world supply chains. For example, we have not considered a multi-stage system with multiple retailers and manufacturers. Extending our results to the multiple retailer case when demand between retailers may be correlated is straightforward.

Assume that there are n retailers and one manufacturer. Retailer k observes customer demand in time $t - 1$, D_{t-1}^k , uses a moving average with p observations to forecast the mean, $\hat{\mu}_t^k$, and standard deviation, $\hat{\sigma}_t^{L,k}$, of the lead time demand, calculates the order-up-to point, y_t^k , and places an order, q_t^k , to the manufacturer. Then, as in the previous section, assuming that excess inventory is returned at no cost, we have,

$$q_t^k = y_t^k - y_{t-1}^k + D_{t-1}^k$$

$$= (1 + L/p)D_{t-1}^k - L/pD_{t-p-1}^k + z(\hat{\sigma}_t^{L,k} - \hat{\sigma}_{t-1}^{L,k}).$$

Here we have assumed that every retailer faces the same lead time, L . The results presented in this section can be modified to include different lead times, L_k , for each retailer.

Given the orders q_t^k placed by each retailer in period t , the manufacturer sees a total order

$$Q_t = \sum_{k=1}^n q_t^k = (1 + L/p) \sum_{k=1}^n D_{t-1}^k - L/p \sum_{k=1}^n D_{t-p-1}^k + z \sum_{k=1}^n (\hat{\sigma}_t^{L,k} - \hat{\sigma}_{t-1}^{L,k}).$$

To quantify the bullwhip effect in this case, we need to determine the variance of the total order seen by the manufacturer relative to the variance of the total demand seen by the retailers. We assume that the demands observed by the retailers are i.i.d. over time, but may be correlated across the retailers. Then, under certain conditions on the symmetry of the joint distribution of demand across retailers, or if $z = 0$, we can state the following lower bound on the variance of the total order seen by the manufacturer relative to the variance of the total customer demand.

Theorem 14.3.4 *If each retailer uses a simple moving average forecast with p demand observations, and an order-up-to inventory policy defined in (14.1), then the variance of the total order, Q , placed by the retailers to the manufacturer, satisfies*

$$\frac{Var(Q)}{Var(\sum_{k=1}^n D^k)} \geq 1 + \frac{2L}{p} + \frac{2L^2}{p^2}. \quad (14.14)$$

The bound is tight when $z = 0$.

It is now clear that the lower bound on the relative increase in variability seen by the manufacturer in the case of multiple retailers is identical to the lower bound on the relative increase in variability seen by the manufacturer in the case of a single retailer, given in (14.8). For a more detailed discussion of the multiple retailer case, see Ryan (1997).

14.4 MANAGERIAL INSIGHTS ON THE BULLWHIP EFFECT

In this section we present some important insights regarding the bullwhip effect that can be derived from the expressions given above for the variance of the orders placed by the retailer relative to the variance of customer demand. In addition, we also present some important insights into the impact of different forecasting methods and different types of demand patterns, e.g., demands with an increasing mean, on the bullwhip effect.

14.4.1 **The bullwhip effect is due, in part, to the need to forecast demand.**

As discussed above, the bullwhip effect is caused, in part, by the need to forecast the mean and standard deviation of demand. In fact, if the retailer follows a simple

order-up-to inventory policy in which she does not update the forecast of the mean and standard deviation of demand in each period, then the target inventory level will remain the same in each period, and there will be no increase in variability, i.e., there will be no bullwhip effect. The results presented in the previous section clearly demonstrate the impact of demand forecasting on the bullwhip effect.

Although this insight seems to suggest that we should eliminate demand forecasting, this clearly is not the answer. Forecasting, of course, provides valuable information to each stage of the supply chain. Without demand forecasting, we would be forced to simply “guess” the expected demand in each period.

14.4.2 Smoother demand forecasts can reduce the bullwhip effect.

As discussed above, if a retailer follows a simple order-up-to inventory policy, such as (14.1), then any forecasting technique which updates the mean and standard deviation of demand in each period will cause the bullwhip effect. We can reduce this effect by using a forecasting technique that produces smooth forecasts, that is, a forecasting technique in which the estimates of the mean and standard deviation of demand do not change dramatically from period to period.

Consider, for example, the moving average forecast, in which we estimate the mean demand as the average of the previous p demand observations. Clearly, when p is small, the estimate of the mean will change more rapidly than when p is large. In other words, larger values of p create smoother forecasts. This implies that the larger p , i.e., the more demand information used to construct the forecast, the lower the bullwhip effect. This is exactly the effect we saw in the previous section. Similarly, for exponential smoothing, small values of α produce smoother forecasts. Therefore, as seen in the previous section, small values of α mean less variability.

14.4.3 With long lead times we must use more demand data to reduce the bullwhip effect.

We have already seen that lead times can add to the bullwhip effect by magnifying the increase in variability due to the demand forecasting. To understand the relationship between the lead time and forecasting technique more fully, note that the expression for the variance of the orders placed by the retailer given in equation (14.7) is an increasing function of the ratio of the lead time to the number of demand observations. In other words, the variability of the orders placed by the retailer increases in the ratio L/p . Similarly, equation (14.10) is an increasing function of $L\alpha$. This fact leads to the following conclusion: a retailer faced with long lead times must use more demand data in her forecasts in order to reduce the bullwhip effect. In other words, when the lead time increases, the retailer must use smoother demand forecasts in order to prevent a further increase in variability.

14.4.4 The magnitude of the increase in variability depends on the forecasting method.

It is now clear that demand forecasting can cause the bullwhip effect. We can also ask whether the **magnitude** of the increase in variability depends on the forecasting method. To answer this question, we compare the magnitude of the increase in variability for the two forecasting methods considered above, moving average and exponential smoothing, for i.i.d. demands and $z = 0$. To do this, we design the forecasts, i.e., choose the parameters p and α , such that both forecasts have the same variance of the forecast errors. It is easily verified, see Nahmias (1993), that if the demands seen by the retailer are i.i.d., then the standard deviation of the forecast errors for exponential smoothing forecasts with smoothing parameter α , is given by

$$\sigma_e = \sigma \sqrt{\frac{2}{2 - \alpha}},$$

while for moving average forecasts, the standard deviation of the forecast errors is

$$\sigma_e = \sigma \sqrt{\frac{p + 1}{p}},$$

where p is the number of observations used in the moving average. If we equate these standard deviations, we obtain

$$\alpha = \frac{2}{p + 1}.$$

Using this relationship between α and p , by equation (14.12), we have

$$\frac{Var(q^{EX})}{\sigma^2} = 1 + \frac{4L}{p + 1} + \frac{4L^2}{p(p + 1)}.$$

On the other hand, when the retailer uses a simple moving average with p observations, we have (14.8). Since

$$\frac{4L}{p + 1} + \frac{4L^2}{p(p + 1)} = \left(\frac{2L}{p} + \frac{2L^2}{p^2}\right) \frac{2p}{p + 1} > \frac{2L}{p} + \frac{2L^2}{p^2},$$

for $p \geq 1$, we see that the increase in variability for simple exponential smoothing is greater than the increase in variability for the simple moving average. In other words, if demands are i.i.d. and $z = 0$, the bullwhip effect will be larger for a retailer using an exponential smoothing forecast than for a retailer using a moving average forecast.

14.4.5 The magnitude of the increase in variability depends on the demand process.

We have already seen that the increase in variability due to demand forecasting depends on the exact nature of the customer demand process. For example, we can see from (14.7) and (14.10) that the magnitude of the increase in variability depends on ρ , the correlation parameter. In particular, we've seen that negatively correlated demands can lead to higher variability than positively correlated demands.

So far we have considered forecasting only the mean and standard deviation of demand under the assumption that the mean demand remains constant over time. Of course, it is frequently the case that the average customer demand changes over time. For example, the average demand may be steadily increasing over time, as in the case of a new product whose market share is growing steadily over time. On the other hand, the average demand for an item may fluctuate over time, as in the case of a seasonal product whose demand may be higher in the summer and lower in the winter. We can also look at the impact of these different types of demand patterns on the bullwhip effect.

As an example, consider the simple supply chain model discussed in the previous section, but now assume that customer demands are independent random variables of the form

$$D_t = \mu + bt + \epsilon_t, \quad (14.17)$$

where μ represents the constant demand level at time $t = 0$, b denotes a linear trend factor, and ϵ_t is the random error term. Assume that the error terms are i.i.d. from a symmetric distribution with mean 0 and variance σ^2 . This implies that the D_t are independent with mean $\mu + bt$ and variance σ^2 . In this case, we can use the standard two parameter exponential smoothing method to forecast customer demand, as described in Hax and Candea (1984, p. 160). As in exponential smoothing with no trend, the estimate of the constant level, μ , is

$$\hat{\mu}_t = \alpha_1 D_{t-1} + (1 - \alpha_1) \hat{\mu}_{t-1}.$$

Similarly, the estimate of the trend parameter, b , is

$$\hat{b}_t = \alpha_2 (\hat{\mu}_t - \hat{\mu}_{t-1}) + (1 - \alpha_2) \hat{b}_{t-1},$$

where $0 < \alpha_1 \leq 1$ and $0 < \alpha_2 \leq 1$ are smoothing constants. Given these estimates, the forecast L period demand is

$$\hat{\mu}_t^L = L\hat{\mu}_t + \left(\frac{L}{\alpha_1} + \frac{L(L-1)}{2} \right) \hat{b}_t.$$

We can repeat the analysis above to obtain the following bound on the variance of the orders placed by a retailer using exponential smoothing to forecast a demand process with a trend, $Var(q_t^{EXT})$:

$$\begin{aligned} \frac{Var(q_t^{EXT})}{\sigma^2} &\geq 1 + 2L\alpha_1 + \frac{2L^2\alpha_1^2}{2 - \alpha_1} \\ &\quad + [2L\alpha_2 + L(L-1)\alpha_1\alpha_2] \times \\ &\quad [1 + \frac{L\alpha_1}{2 - \alpha_1} + \frac{(\alpha_1 + \alpha_2)(L\alpha_1 + L\alpha_2 + \frac{L(L-1)}{2}\alpha_1\alpha_2)}{(2 - \alpha_1)(\alpha_1 + \alpha_2 - \alpha_1\alpha_2)}]. \end{aligned}$$

From this expression, we can see that, for a retailer using exponential smoothing with $\alpha = \alpha_1$, any choice of α_2 , and $z = 0$, the increase in variability for the linear trend demand model is greater than the increase in variability, given by (14.12), for the i.i.d. demand model.

We also note that the increase in variability does not depend on the true value of the trend parameter, b , but rather on the smoothing parameter for the linear trend, α_2 . In other words, the additional increase in variability is not due to the magnitude of the linear trend, but to the need to estimate an additional parameter, b , when forecasting a demand process with a linear trend. In particular, consider a retailer who faces i.i.d. demands, but who believes she is facing a demand process with a linear trend, and therefore uses an exponential smoothing forecast with a linear trend. The orders placed by this retailer will be more variable than those placed by the same retailer using a simple (without trend) exponential smoothing forecast.

14.5 THE IMPACT OF CENTRALIZED INFORMATION ON THE BULLWHIP EFFECT

One of the most frequent suggestions for reducing the bullwhip effect is to centralize demand information within a supply chain, that is, to provide each stage of the supply chain with complete information on the actual customer demand. For example, Lee, Padmanabhan, and Whang (1997a) suggest that “one remedy is to make demand data at a downstream site available to the upstream site.” To understand why centralized demand information can reduce the bullwhip effect, note that if demand information is centralized, each stage of the supply chain can use the actual customer demand data to create more accurate (less variable) forecasts, rather than relying on the orders received from the previous stage, which, as we have seen, can be significantly more variable than the actual customer demand.

In this section we consider the value of sharing customer demand information within a supply chain. To determine the impact of centralized demand information on the bullwhip effect, we will consider two types of supply chain, with and without centralized demand information, both of which are described below.

14.5.1 Centralized Demand Information

Consider a multi-stage supply chain in which the first stage (i.e., the retailer) shares all demand information with each of the subsequent stages. In other words, in each period the retailer provides every stage of the supply chain with complete information on customer demand. Assume that all stages of the supply chain use a moving average forecast with p observations to estimate the mean demand, and that the demands seen by the retailer are i.i.d.. Therefore, since each stage has complete information on customer demand, each stage will use the same estimate of the mean demand, $\hat{\mu}_t = \sum_{i=1}^p D_{t-i}/p$.

Finally, assume that each stage follows an order-up-to inventory policy in which the order-up-to point in each period is just the forecast demand over a fixed lead time. That is, assume that each stage, k , follows an order-up-to policy where the order-up-to point is of the form

$$y_t^k = L_k \hat{\mu}_t.$$

Note that this inventory policy is just a special case of the policy defined in (14.1) with $z = 0$. Frequently, when a policy of this form is used in practice, an inflated value of the lead time L_k is used, with the excess inventory representing safety stock.

For example, a retailer facing an order lead time of three weeks may choose to keep inventory equal to four weeks of forecast demand, with the extra week of inventory representing his safety stock. In our experience, policies of this form are often used in practice. Indeed, we have recently collaborated with a major U.S. retail company that uses a policy of this form. See also Johnson *et al.* (1995).

In this case, if we perform an analysis similar to that presented above, we have the following expression for the variance of the orders placed by stage k , q^k , relative to the variance of customer demand.

$$\frac{Var(q^k)}{Var(D)} = 1 + \frac{2(\sum_{i=1}^k L_i)}{p} + \frac{2(\sum_{i=1}^k L_i)^2}{p^2} \quad \forall k. \quad (14.22)$$

Notice that this expression is quite similar to the single stage bound presented in (14.8), but with the single stage lead time, L , replaced by the echelon lead time, $\sum_{i=1}^k L_i$. This result also provides a lower bound on the increase in variability when the retailer uses an inventory policy in which z is not equal to zero, as defined in (14.1).

In many supply chains where customer demand information is shared between stages an echelon inventory policy is used. Consider an echelon inventory policy where the order-up-to point is of the form

$$y_t^k = \left(\sum_{i=1}^k L_i \right) \hat{\mu}_t,$$

where $\sum_{i=1}^k L_i$ is the echelon lead time. The expression given in (14.22) for the variance of the orders placed by each stage of the supply chain also holds in this case. In other words, if demand information is centralized, then the increase in variability seen by each stage of the supply chain is the same whether the supply chain follows an echelon inventory policy or not.

Finally, notice that (14.22) demonstrates that even when (i) all demand information is centralized, (ii) every stage of the supply chain uses the same forecasting technique, and (iii) every stage of the supply chain uses an echelon inventory policy, we will still see an increase in variability at each stage of the supply chain. In other words, by centralizing customer demand information and coordinating inventory control we have not completely eliminated the bullwhip effect.

14.5.2 Decentralized Demand Information

Next, consider a supply chain similar to the one just analyzed, but without centralized customer demand information. In this case, the retailer does not provide the upstream stages with any customer demand information. Therefore, each stage determines its forecast demand based on the orders placed by the previous stage, not based on actual customer demand. Assume that each stage of the supply chain follows an order-up-to policy of the form

$$y_t^k = L_k \hat{\mu}_t^{(k)},$$

where L_k is the lead time between stages k and $k+1$,

$$\hat{\mu}_t^{(1)} = \frac{\sum_{i=1}^p D_{t-i}}{p},$$

and

$$\hat{\mu}_t^{(k)} = \frac{\sum_{j=0}^{p-1} q_{t-j}^{k-1}}{p} \quad \forall k \geq 2,$$

where q_t^k is the order placed by stage k in period t .

In this case, we have the following lower bound on the variance of the orders placed by each stage of the supply chain:

$$\frac{Var(q^k)}{Var(D)} \geq \prod_{i=1}^k \left(1 + \frac{2L_i}{p} + \frac{2L_i^2}{p^2} \right) \quad \forall k. \quad (14.23)$$

Note that this expression for the variance of the orders placed by the k th stage of the supply chain is similar to the expression for the variance of the orders placed by the retailer given in (14.8), but now the variance increases multiplicatively at each stage of the supply chain. Also, again note that the variance of the orders becomes larger as we move up the supply chain, so that the orders placed by the manufacturer are more variable than the orders placed by the retailer.

14.5.3 Managerial Insights on the Value of Centralized Information

We are interested in determining the impact of centralized customer demand information on the bullwhip effect. To do this, we must compare the increase in variability at each stage of the supply chain for the centralized and decentralized systems. We have already seen that for either type of supply chain, i.e., centralized or decentralized, the variance of the order quantities becomes larger as we move up the supply chain, so that the orders placed by the manufacturer are more variable than the orders placed by the retailer, and so on. The difference in the two types of supply chains is in terms of how much the variability grows as we move from stage to stage.

The results above indicate that, for supply chains with centralized information, the increase in variability at each stage is an additive function of the lead time and the lead time squared, while for supply chains without centralized information, the lower bound on the increase in variability at each stage is multiplicative. In other words, a decentralized supply chain, in which only the retailer knows the customer demand, can lead to significantly higher variability than a centralized supply chain, in which customer demand information is available at each stage of the supply chain, particularly when lead times are large. So we conclude that **centralizing customer demand information can significantly reduce the bullwhip effect**.

It is clear that by sharing demand information with each stage of the supply chain we can significantly reduce the bullwhip effect. This reduction in variability is due to the fact that when demand information is centralized, each stage of the supply chain can use the actual customer demand data to estimate the average demand. On the other hand, when demand information is not shared, each stage must use the orders placed by the previous stage to estimate the average demand. As we have already seen, these orders are more variable than the actual customer demand data, and thus the forecasts created using these orders are more variable, leading to even more variable orders.

Figure 4 presents the variability in orders placed by the manufacturer (i.e., stage 2 in the supply chain) for the case of centralized (dashed lines) and decentralized

(dotted lines) demand information when $L = 1$. In each case, the figure presents the increase in variability for $z = 0$ (from (14.22) in the centralized case and (14.23) in the decentralized case) and for $z = 2$ (based on simulation results). It also presents, as a reference, the increase in variability for the first stage (solid lines) for both $z = 0$ (from (14.8)) and $z = 2$ (simulation results).

Figure 5 presents the variability of the orders placed by stages 1, 2 and 3 of the supply chain, when $z = 0$ and $L_1 = L_2 = L_3 = 3$, for centralized (dashed lines) and decentralized supply chains (dotted lines). This figure, when compared with Figure 4, indicates that in multi-stage supply chains, the lead time can have a significant impact on the increase in variability. In addition, this figure indicates that the **difference** between the variance in the centralized and decentralized supply chains increases as we move up the supply chain, i.e., as we move from the second to the third stage of the supply chain.

Finally, both Figures 4 and 5 demonstrate that centralizing customer demand information can reduce variability. In addition, we see that for *both* the centralized and the decentralized systems there is an increase in variability when we move from the first stage to the second stage to the third stage. This figure reinforces an important insight: **the bullwhip effect exists even when demand information is completely centralized and all stages of the supply chain use the same forecasting technique and inventory policy**. So we conclude that centralizing customer demand information can significantly reduce, but will not eliminate, the bullwhip effect.

14.6 METHODS FOR REDUCING THE IMPACT OF THE BULLWHIP EFFECT

The research on identifying and quantifying the causes of the bullwhip effect has led to a number of suggestions for reducing the bullwhip effect or for eliminating its impact. These suggestions, as discussed in Lee, Padmanabhan and Whang (1997 a, b), include reducing uncertainty, reducing the variability of the customer demand process, reducing lead times and engaging in strategic partnerships. These suggestions, and how they relate to the new insights presented above, are discussed briefly below.

14.6.1 Reducing Uncertainty

As mentioned above, one of the most frequent suggestions for reducing or eliminating the bullwhip effect is to reduce uncertainty throughout the supply chain by centralizing customer demand information. The results presented in the previous section demonstrate that centralizing demand information can in fact reduce (but will not eliminate) the bullwhip effect.

Lee, Padmanabhan and Whang (1997a) also point out, however, that even if each stage uses the same demand data, they may use different forecasting methods and different buying practices or ordering policies, both of which may contribute to the bullwhip effect. The results presented in the previous section indicate that even when each stage uses the same demand data, the same forecasting method and the same ordering policy, the bullwhip effect will continue to exist.

14.6.2 Reducing Variability

We can also reduce the bullwhip effect by reducing the variability inherent in the customer demand process. For example, if we can reduce the variability of the customer demand seen by the retailer, then even if the bullwhip effect occurs, the variability of the demand seen by the manufacturer will also be reduced.

We can reduce the variability of customer demand through, for example, the use of an “every day low pricing” strategy, or EDLP. When a retailer uses EDLP, he offers a product at a single consistent price, rather than offering a regular price with periodic price promotions. By eliminating price promotions, a retailer can eliminate many of the dramatic shifts in demand that occur along with these promotions. Therefore, every day low pricing strategies can lead to much more stable, i.e., less variable, customer demand patterns.

14.6.3 Leadtime Reduction

The results presented in this paper clearly indicate that lead times serve to magnify the increase in variability due to demand forecasting and demonstrate the dramatic effect that lead times can have on the variability at each stage of the supply chain. Therefore, it is clear that lead time reduction can significantly reduce the bullwhip effect throughout a supply chain.

The results presented here also demonstrate the relationship between the lead time and the forecasting technique. In particular, we have demonstrated that with longer lead times a retailer must use more demand data (smoother forecasts) in order to reduce the bullwhip effect.

14.6.4 Strategic Partnerships

Finally, we can also eliminate the bullwhip effect by engaging in any of a number of strategic partnerships. These strategic partnerships change the way information is shared and inventory is managed within the supply chain, and can therefore reduce or eliminate the impact of the bullwhip effect. For example, in vendor managed inventory (VMI), the manufacturer manages the inventory of his product at the retailer, determining for himself how much inventory to keep on hand and how much to ship to the retailer in every period. Therefore, the manufacturer does not rely on the orders placed by the retailer, and thus avoids the bullwhip effect entirely.

Other types of partnerships can be designed to reduce the bullwhip effect. For example, the results presented in this paper indicate that centralizing demand information can dramatically reduce the variability seen by the upstream stages in a supply chain. Therefore, it is clear that these upstream stages would benefit from a strategic partnership which provides an incentive for the retailer to make customer demand data available to the rest of the supply chain.

References

- Baganha, M. and M. Cohen, "The Stabilizing Effect of Inventory in Supply Chains," To appear in *Operations Research*, 1995.
- Caplin, A.S., "The Variability of Aggregate Demand with (S,s) Inventory Policies," *Econometrica*, 53 (1985), 1396–1409.
- Chen, F., J. K. Ryan and D. Simchi-Levi, "The Impact of Exponential Smoothing Forecasts on the Bullwhip Effect," Working Paper, Northwestern University, 1998.
- Chen, F., Z. Drezner, J. K. Ryan, and D. Simchi-Levi, "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Leadtimes and Information," Working Paper, Northwestern University, 1998.
- Johnson, M.E., H.L. Lee, T. Davis and R. Hall (1995), Expressions for Item Fill Rates in Periodic Inventory Systems. *Naval Research Logistics*, 42, pp. 39–56.
- Kahn, J., "Inventories and the Volatility of Production," *The American Economic Review*, 77 (1987), 667–679.
- Kaminsky, P. and D. Simchi-Levi, A New Computerized Beer Game: Teaching the Value of Integrated Supply Chain Management. To appear in the book *Supply Chain and Technology Management*. Hau Lee and Shu Ming Ng, eds., the Production and Operations Management Society, (1996).
- Lee, H., P. Padmanabhan and S. Whang, "The Bullwhip Effect in Supply Chains," *Sloan Management Review*, 38 (1997a), 93–102.
- Lee, H., P. Padmanabhan and S. Whang, "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science*, 43 (1997b), 546–58.
- Ryan, J. K., "Analysis of Inventory Models with Limited Demand Information," Ph.D. Dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, 1997.
- Sterman, J. D., "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment," *Management Science*, 35 (1989), 321–339.

Figure 1: $\text{Var}(q)/\text{Var}(D)$ for $\rho=0.5$ and 0.9 , $L=1$, $z=2$.

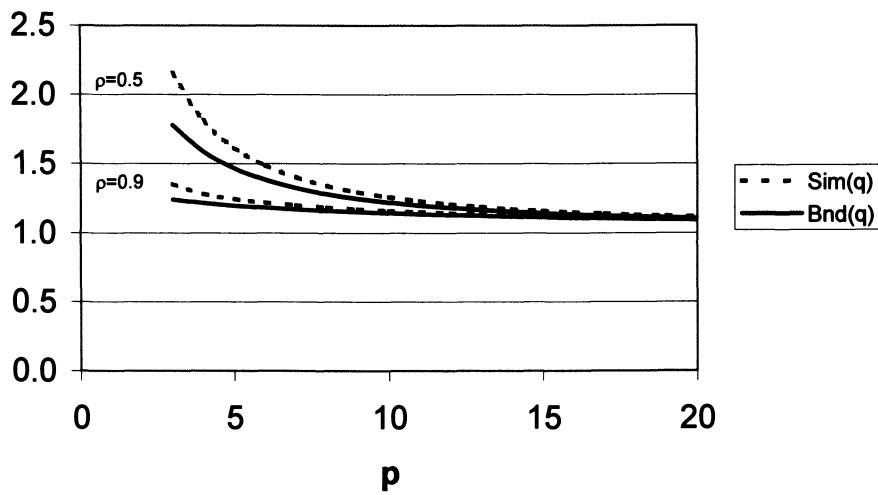
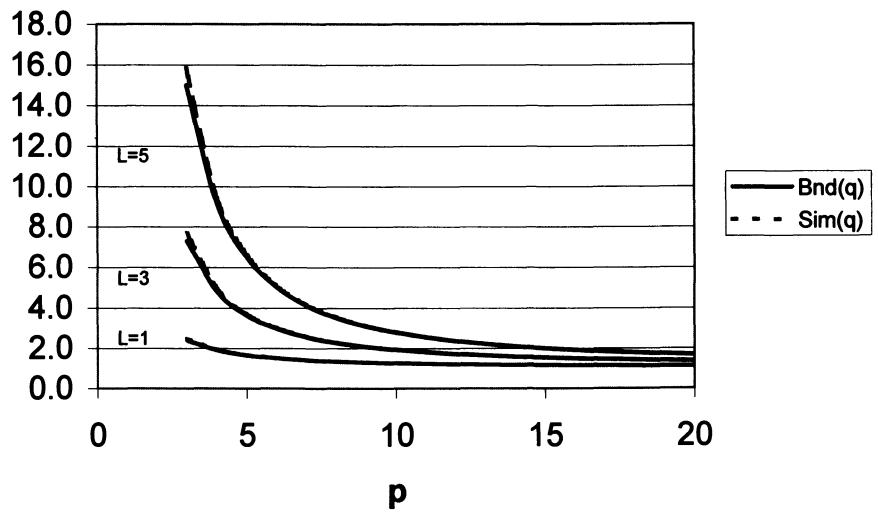
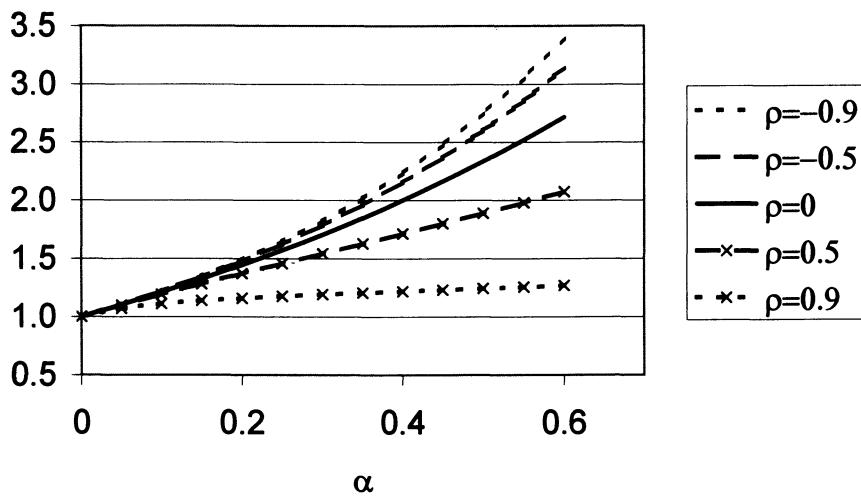


Figure 2: $\text{Var}(q)/\text{Var}(D)$ for $L=1, 2$ and 3 , $\rho = 0$, $z=2$.



**Figure 3: $\text{Var}(q)/\text{Var}(D)$ for various values of ρ ,
 $L=1, z=0$.**



**Figure 4: $\text{Var}(q^k)/\text{Var}(D)$ Centralized vs. Decentralized for $k=1$ and
 $2, L=1$.**

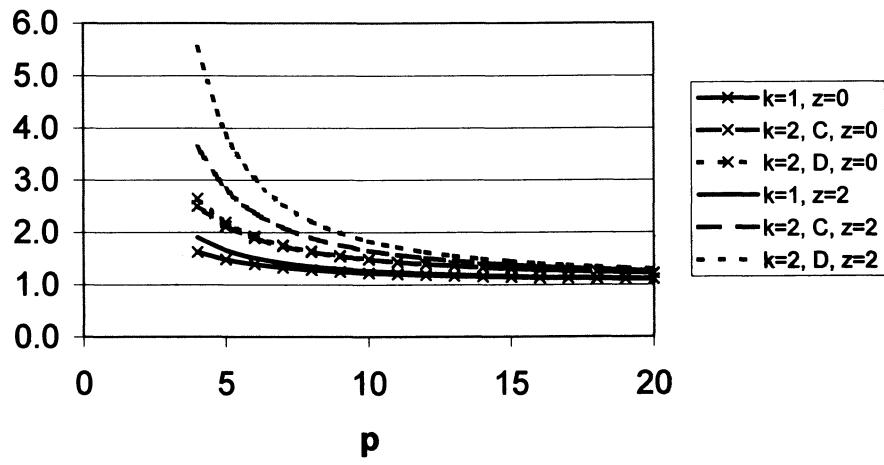


Figure 5: $\text{Var}(q^k)/\text{Var}(D)$ Centralized vs. Decentralized for $k=1, 2,$
 $\text{and } 3, L=3, z=0.$

