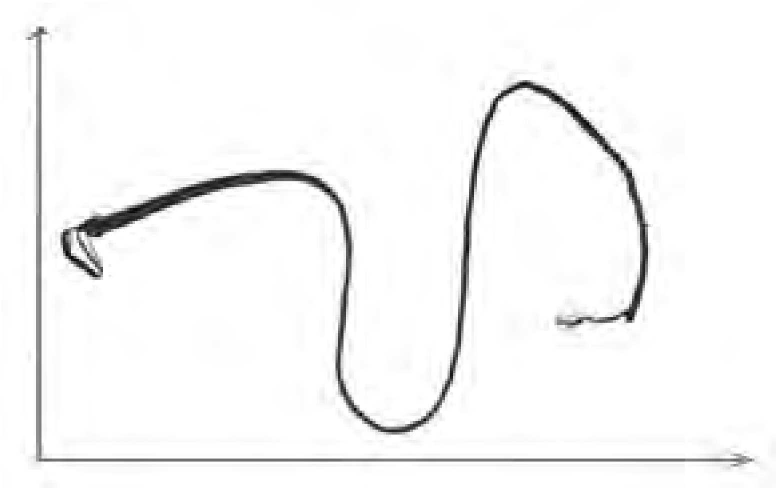
VU University Amsterdam

Research Paper Business Analytics

The Bullwhip Effect: Analysis of the Causes and Remedies

Author:
Jonathan Moll

Supervisor:
Rene Bekker



VU University Amsterdam
Faculty of Sciences
Boelelaan 1081a
1081 HV Amsterdam

July 2013

Preface

This research paper is a compulsory part of the Master's program Business Analytics at the VU University Amsterdam. The student should do research on a business related problem, with a strong mathematical or computer science component. The research could be based on literature, but may also be extended with own research.

This research paper presents an elaborate literature overview of the bullwhip effect, extended with computer simulations to get more insight in the processes that cause the effect. The bullwhip effect is a phenomenon that refers to a trend of larger and larger swings in inventory, as one moves upstream in a supply chain. Since a long crackling whip is a suitable metaphor for the amplifying and oscillating swings in inventory, the phenomenon is called the bullwhip effect.

I would like to thank Rene Bekker for the time and effort he spend supervising me.

Jonathan Moll
July 2013

Summary

The bullwhip effect is a problem in supply chains that refers to amplifying swings in inventory as one moves upstream along a supply chain (further away from the customer). The aim of this paper is to give an elaborate literature overview, supplemented with computer simulations to make the problem more intelligible.

In Chapter 2 we first describe the bullwhip effect based on the results of Sterman's beer game [22]. The beer game simulates a supply chain consisting of a beer retailer, wholesaler, distributor and brewery. The game is played by four players who make independent inventory decisions without consultation with other players. Typical game results clearly demonstrate the amplification effects.

Secondly, we show how the bullwhip effect could be measured and present the simple supply chain model of Chen et al. [6]. To quantify the size of the bullwhip effect, Chen et al. propose the order rate variance ratio, which is given by the variance of the orders, divided by the variance of the demand. If the ratio is greater than one, the bullwhip effect exists.

Chen et al. assume that the demand follows an autoregressive model of the first degree, an AR(1) model. They further assume that all members of the supply chain use a simple order-up-to policy, where the mean demand and the average forecast error are estimated with moving average forecasts. Given these assumptions, they derive a lower bound for the order rate variance ratio. The ratio is always greater than one, which implies the bullwhip effect.

To make the problem more intelligible we carry out two simulation experiments. We first simulate a simple supply chain using exponential smoothing forecasting, and then simulate the model of Chen et al. The simulation results clearly show the amplifying swings in demand and inventory as one moves upstream the supply chain. It also shows the high levels of safety stock that are needed to compensate for the large swings in inventory. The order rate variance ratio grows fast if the lead time between ordering and receiving grows. The impact of the degree of serial correlation of the demand is also significant. If the correlation parameter increases, then the order rate variance ratio will decrease.

Chapter 3 is concerned with the causes of the bullwhip effect. Forrester [12] and Sterman [22] ascribe behavioural causes to the bullwhip effect, where Lee et al. [17] [18] suggest that the bullwhip effect also occurs due to operational causes, even if the members of the supply chain behave perfectly rational. Lee et al. define four operational causes: Demand signal processing, rationing and shortage gaming, order batching and price fluctuations. Lee et al. [18] also developed mathematical models to proof that the four causes they identified indeed lead to the bullwhip effect. The proofs show that all four causes lead to a situation where the variance of orders is greater than the variance of demand. This implies that the causes lead to the bullwhip effect.

In Chapter 4 we discuss proposed remedies to the bullwhip effect. The most obvious remedy to counter demand signal processing is collaboration, but it generally does not completely eliminate the problem. Another remedy to damp the effect of demand signal processing is replenishment smoothing. With replenishment smoothing, a company does not try to recover the gap between the available inventory and the order-up-to point in one period, but tries to smoothly recover the gap in more than one order period. Finally, operational efficiency is also important to counter the effects of demand signal processing. Operational efficiency leads to shorter lead times which have a negative impact on the bullwhip effect. An effective way to improve operational efficiency is with just-in-time management (JIT).
A way to counter the effects of order batching is the use of electronic data interchange (EDI) and third party logistic providers (3PL). Price fluctuations can be countered by a every-day-low-price strategy (EDLP). With EDLP companies offer a constant (low) price, this leads to more stability in orders. Finally, to prevent rationing and shortage gaming, members of the supply chain could make stricter rules to prevent retailers from overordering in times of shortage and cancelling in times of surplus.

Contents

Chapter 1

Introduction

A key issue in supply chain management is the control of inventory. Members of a supply chain adopt policies and operating procedures to optimize their use of inventory such that it minimizes their investments in inventory while keeping high customer service levels. Uncertainty is inherent to every supply chain through factors such as variability in demand, lead times, breakdowns of machines and local politics. Because uncertainty is given, companies often carry an inventory buffer called safety stock. An important observation in supply chain management is that small variations in demand from customers result in increasingly large variations in demand as one moves up the supply chain (further away from the customer). This phenomenon is known as *the bullwhip effect*. The bullwhip effect can lead to major inefficiencies such as excessive inventory investments, high levels of safety stock, misguided capacity plans, inactive transportation and poor customer service.

The bullwhip effect has been noted across a range of academic disciplines, which assigned various causes and proposed differing remedies for the problem. The effect of demand amplification upstream in a supply chain was first described in 1961 by Forrester [12], who used computer simulation to show the existence of demand amplification and demonstrated its effects. Forrester's fundamental explanation was that it is primarily a function of decision making in response to variability in incoming demand. Forrester's remedy lay in understanding the system as a whole, and modelling that system with specific simulation models, so that managers can determine appropriate action.
Perhaps the most well-known description of the bullwhip effect has come from Sterman [22] in 1989. Sterman illustrates the bullwhip effect with the "beer distribution game" that simulates a supply chain consisting of a beer retailer, wholesaler, distributor and brewery. The game is played by four players who make independent inventory decisions without consultation with other players. The individual players only rely on the orders

of the neighbouring player. Again, the amplification effects were clearly demonstrated. Sterman interprets the phenomenon as a consequence of players' systematic irrational behaviour, or "misperceptions of feedback".
In contradiction to Forrester [12] and Sterman [22], who ascribed the bullwhip effect to irrational behaviour of managers, Lee et al. [17] identify in 1997 four operational causes of the bullwhip effect: Demand signal processing, rationing gaming, order batching, and price discounting. These four high level causes have become the standard for this phenomenon. In the same year, Lee et al. [18] employ mathematical models to demonstrate how these causes can lead, even with rational decision making, to the amplification of demand. Since these papers, much research has been done on the impact of different methods to counter the causes of the bullwhip effect. Some examples are Chen et al. [6], Cachon et al. [3], Disney et al. [8], Canella et al. [4].

This first aim of this paper is to give an extensive literature overview of the bullwhip effect, its causes and proposed remedies. Secondly, it illustrates the bullwhip effect and the impact of different parameters on the bullwhip effect in simulation experiments.
In Chapter 2 we first describe the bullwhip effect based on the results of Sterman's beer game [22]. Then we show how the bullwhip effect can be measured and present the simple supply chain model of Chen et al. [6]. Finally, in the last section of Chapter 2, we carry out simulation experiments to gain more insight in the process that causes the bullwhip effect. Chapter 3 is concerned with the causes of the bullwhip effect. First we describe the behavioural causes advocated by Forrester [12] and Sterman [22]. Then we give an extensive description and present the mathematical proofs for the operational causes identified by Lee et al. [17] [18]. In the last chapter we describe proposed remedies to counter the four operational causes identified by Lee et al. [17]. We conclude with a short discussion about the literature.

Chapter 2

The Bullwhip Effect

Consider multiple companies operating in a serial supply chain, each of whom orders from its immediate upstream neighbour. Research indicates that small variations in demand from customers result in increasingly large variations as demand is transmitted upstream along the supply chain. In this chapter we first give a description of this phenomenon on the basis of the "beer distribution game" from Sterman [22]. Secondly, we give an introduction to the measurement and quantitative analysis of the bullwhip effect on the basis of a simple supply chain model, developed by Chen et al. [6]. In the last section we do simulation experiments to gain more insight in the process that causes the bullwhip effect and to study the impact of different parameters on the size of the effect.

2.1 Description of the Phenomenon

Probably the most well known demonstration of the bullwhip effect was carried out by Sterman [22] with the well known "beer distribution game". The game is a role-playing simulation that portrays the supply chain of beer. It consists of four sectors: retailer, wholesaler, distributor and factory. The game is played in teams of four persons and each person manages one sector. Each week a card is drawn that represents customer demand. The retailer ships the beer requested out of its inventory and orders new beer from the wholesaler, who ships the beer requested out of its inventory and orders beer from the distributor, who orders and receives beer from the factory. So there is a downstream flow of physical goods and an upstream flow of demand information, as displayed in Figure 2.1. At each stage there are order and receiving delays that represent the time required to receive, process, ship, and deliver orders. The objective is to minimize total company costs that consists of inventory holding costs and stockout costs.

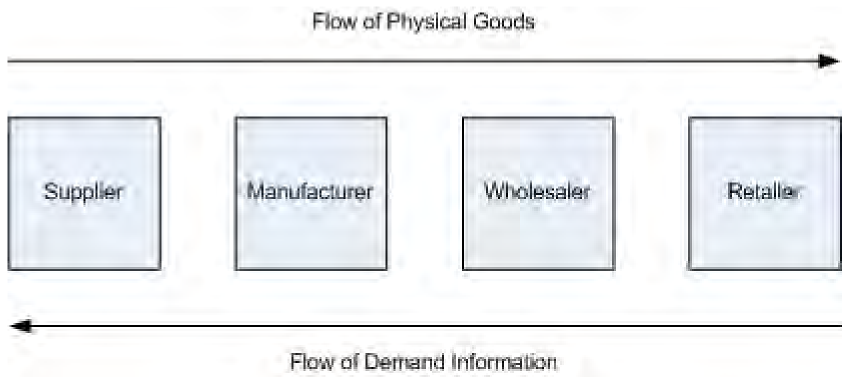


FIGURE 2.1: From Metters [19]. Flow of goods and information in the beergame.

The game has been played all over the world by thousands of people ranging from high school students to chief executive officers and government officials. All of these trials show the following results. Orders and inventory are both subject to instability and oscillation. In almost all cases, the inventory levels of the retailer decline, followed by a decline in inventory of the wholesaler, distributor and factory. As inventory falls, players tend to increase their orders. As additional beer is brewed, inventory levels grow and in many cases overshoot the desired level. As a reaction, orders fall of rapidly. In addition, the amplitude and variance of orders increases as one moves from retailer to factory and the order rate tends to peak later at each stage. Figure 2.2 shows the effects in four typical trials. The oscillation, amplification and phase lag effects are clearly visible.

The bullwhip effect is not solely observed in theory and simulations, it has been recognized in many markets. Procter & Gamble (P&G) found strange order patterns for diapers. The sales at retail stores were fluctuating, but the fluctuations were relatively small compared to the large variability of orders placed by the distributor. The distributors orders to their suppliers were even larger. This did not make sense, because the babies consumed diapers at a steady rate, whereas the variations in demand were amplified as they moved up the supply chain. P&G called this phenomenon the bullwhip effect. The same phenomenon was observed by executives of Hewlett-Packard in the supply chain for laser printers [17]. Also many examples can be found in the grocery industry [1], and automotive industry [2].

2.2 Measurement and Quantitative Analysis

Measurement and quantitative analysis of the bullwhip effect is of great importance for both theoretical and empirical purposes. The most common measure to quantify the bullwhip effect is the order rate variance ratio. This measure was proposed by Chen et

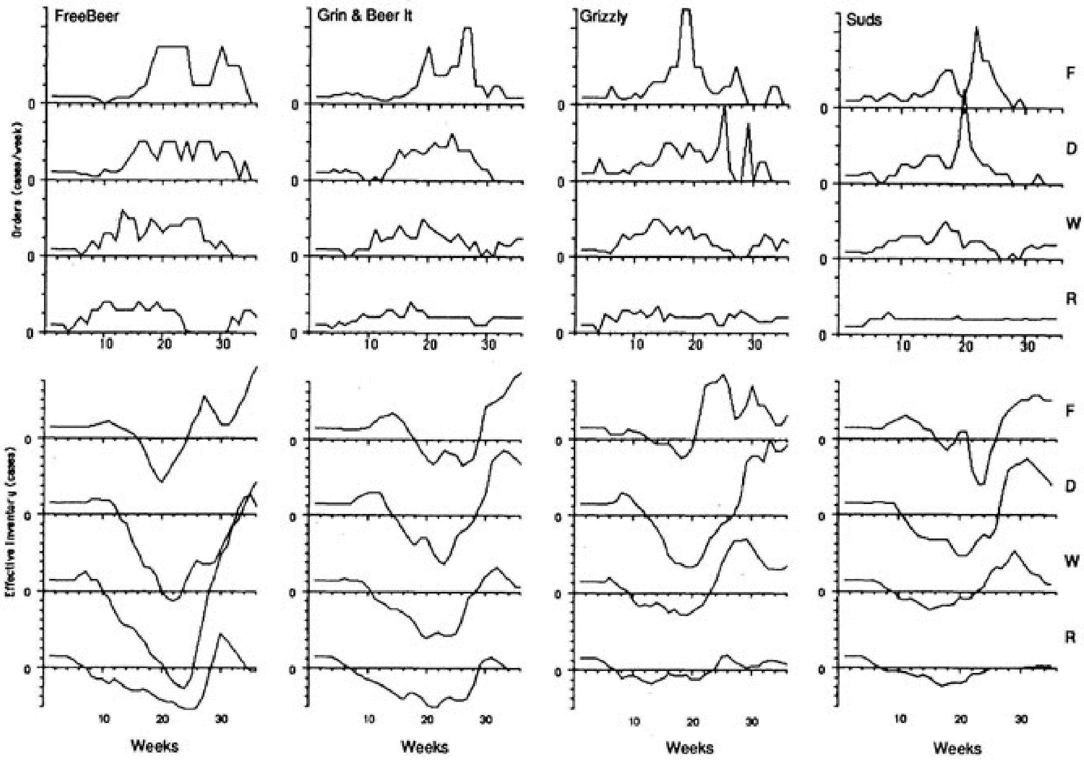


FIGURE 2.2: From Sterman [22]. Top: orders. Bottom: inventory. From bottom to top: retailer, wholesaler, distributor, factory.

al. [6] and is defined as the ratio of the variance of placed orders and observed customer demand. The larger the order rate variance ratio, the stronger the bullwhip effect. In the same paper Chen et al. use a simple supply chain model to extract a lower bound on the order rate variance ratio at each stage of the supply chain. A description of this model and a summary of the results are given below.

2.2.1 A Simple Supply Chain Model

Consider a simple supply chain in which in each period t, a single retailer observes its inventory level and places an order q_t to a single manufacturer. The observed demand at the end of period t is denoted by D_t. There is a fixed lead time of L periods between the time an order is placed by the retailer and the time the order is received from the manufacturer. The observed demand is assumed to be of the form

$$D_t = \mu + \rho D_{t-1} + \epsilon_t, \tag{2.1}$$

where μ is a non-negative constant, ρ is a correlation parameter with $|\rho| < 1$, and the error terms, ϵ_t, are independent and identically distributed (i.i.d) from a symmetric distribution with mean 0 and variance σ^2. The demand process (2.1) is known as a first-order autoregressive model (AR(1) model). The expectation and variance of D_t can easily be derived.

$$E[D_t] = E[\mu] + \rho E[D_{t-1}] + E[\epsilon_t] = \mu + \rho\mu,$$

so

$$E[D_t] = \frac{\mu}{1-\rho}. \tag{2.2}$$

For the variance it holds that

$$Var(D_t) = Var(\mu) + \rho^2 Var(D_{t-1}) + Var(\epsilon_t) = \rho^2 Var(D_{t-1}) + \sigma^2.$$

Because $|\rho| < 1$, $Var(D_t)$ becomes independent of t, and hence

$$Var(D_t) = \frac{\sigma^2}{(1-\rho)^2}. \tag{2.3}$$

The Inventory Policy and Forecasting Technique

Chen et al. [6] assume that the retailer follows a simple *order-up-to inventory policy* in which the order-up-to point y_t is estimated from the observed demand as

$$y_t = \hat{D}_t^L + z\hat{\sigma}_{et}^L, \tag{2.4}$$

where $\hat{D}_t^L$ is an estimate of the mean lead time demand in period t, $\hat{\sigma}_{et}^L$ is an estimate of the standard deviation of the L period forecast error in period t, and z is a constant chosen to meet a desired service level.

To estimate $\hat{D}_t^L$ and $\hat{\sigma}_{et}^L$ Chen et al. assume that the retailer uses a simple moving average based on the demand observation from the previous p periods. That is,

$$\hat{D}_t^L = L\left(\frac{\sum_{i=1}^p D_{t-i}}{p}\right), \tag{2.5}$$

and

$$\hat{\sigma}_{et}^L = C_{L,\rho}\sqrt{\frac{\sum_{i=1}^p (e_{t-i})^2}{p}}, \tag{2.6}$$

where e_{t-i} is the forecast error in period $t-i$ and $C_{L,\rho}$ is a constant function of L, ρ and p. For details about this constant, the reader is referred to Ryan [20]. The use of a simple moving average for the estimaters $\hat{D}_t^L$ and $\hat{\sigma}_{et}^L$ is in general not optimal. However, this method is often used in practice and is therefore a realistic choice.

Quantifying the Bullwhip Effect

To quantify the bullwhip effect, Chen et al. [6] determine 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, q_t is written as

$$q_t = y_t - y_{t-1} + D_{t-1}. \tag{2.7}$$

Note that q_t can be negative, but in that case it is assumed that this excess inventory is returned without cost. Given (2.5) and (2.6), q_t can be written as

$$\begin{aligned} q_t &= \hat{D}_t^L - \hat{D}_{t-1}^L + z(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L) + D_{t-1} \\ &= L\left(\frac{D_{t-1} - D_{t-p-1}}{p}\right) + D_{t-1} + z(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L) \\ &= (1 + L/p)D_{t-1} - (L/p)D_{t-p-1} + z(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L). \end{aligned}$$

Chen et al. [6] evaluate the variance of q_t as follows. First note that D_{t-1} is independent of $\hat{\sigma}_{e,t-1}^L$ and D_{t-p-1} is independent of $\hat{\sigma}_{et}^L$. Since $Cov(D_{t-p-1}, \hat{\sigma}_{e,t-1}^L) = Cov(D_{t-1}, \hat{\sigma}_{et}^L)$, we have

$$\begin{aligned} Var(q_t) &= \left(1 + \frac{L}{p}\right)^2 Var(D_{t-1}) - 2\left(\frac{L}{p}\right)\left(1 + \frac{L}{p}\right) Cov(D_{t-1}, D_{t-p-1}) \\ &\quad + \left(\frac{L}{p}\right)^2 Var(D_{t-p-1}) + z^2 Var(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L) \\ &\quad + 2z\left(1 + \frac{2L}{p}\right) Cov(D_{t-1}, \hat{\sigma}_{et}^L) \\ &= \left(1 + \frac{2L}{p} + \frac{2L^2}{p^2}\right) Var(D) - \left(\frac{2L}{p} + \frac{2L^2}{p^2}\right) + \rho^p Var(D) \\ &\quad + 2z\left(1 + \frac{2L}{p}\right) Cov(D_{t-1}, \hat{\sigma}_{et}^L) + z^2 Var(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L) \\ &= \left[1 + \left(\frac{2L}{p} + \frac{2L^2}{p^2}\right)(1 - \rho^p)\right] Var(D) + 2z\left(1 + \frac{2L}{p}\right) Cov(D_{t-1}, \hat{\sigma}_{et}^L) \\ &\quad + z^2 Var(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L), \end{aligned} \tag{2.8}$$

where the second equation follows from $Cov(D_{t-1}, D_{t-p-1}) = [\rho^p/(1-\rho^2)]\sigma^2$ and $Var(D_t) = \sigma^2/(1 - \rho^2)$. To further evaluate $Var(q_t)$ the following lemma was used. A proof of this lemma is given by Ryan [20].

Lemma 2.1. *Assume the customer demands seen by the retailer are random variables of the form given in (2.1) where the error terms, ϵ_t, are i.i.d from a symmetric distribution with mean 0 and variance σ^2. Let the estimate of the standard deviation of the L period forecast error be $\hat{\sigma}_{et}^L$ as defined in (2.6). Then*

$$Cov(D_{t-i}, \hat{\sigma}_{et}^L) = 0 \qquad \textit{for all } i = 1, \ldots, p.$$

By rewriting Equation (2.8) using Lemma 2.1, Chen et al. [6] found the following lower bound on the increase in variability from the retailer to the manufacturer.

$$\frac{Var(q)}{Var(D)} \geq 1 + \left(\frac{2L}{p} + \frac{2L^2}{p^2}\right)(1 - \rho^p). \tag{2.9}$$

When $z = 0$, the bound is tight.

Interpretation of the Results

Several observations can be made from (2.9). First, since $L \geq 1$, the increase in variability of orders from the retailer to the manufacturer is greater than one, and thus the bullwhip effect occurs. If $0 \leq \rho \leq 1$, the ratio is a decreasing function of p, a decreasing or alternating function of ρ, and an increasing function of L. The number of observations p used to estimate the mean and variance of demand have a significant impact. Note that if p becomes large, the lower bound becomes one. However, because (2.9) gives only a lower bound, it is not necessarily true that the bullwhip effect will not occur with perfectly accurate estimates.

A remarkable fact is that if $\rho < 0$, then ρ^p is alternating, which suggests that the increase in variability will be larger for odd values of p than for even values of p. This is logical, because if ρ is negative, the demand is alternating and the estimation of the mean and variance will be more accurate after an even number of observations. For example, suppose that if you estimate from three data points, you estimate from two high numbers and one low number. This will lead to a high estimate for the next number, while it will be low because of the negative correlation. In contrast, if you would estimate from four data points, you will estimate from two high numbers and two low numbers. Your estimate will be in the middle, which will give a more accurate estimate than with three data points.

Chen et al. [6] do not say anything about the impact of the customer service level parameter z on the order rate variance ratio. The customer service level z does not appear in the lower bound. It only appears in the exact equation for the order rate variance ratio, which can easily be derived from (2.8), using Lemma 2.1.

$$\frac{Var(q)}{Var(D)} = 1 + \left(\frac{2L}{p} + \frac{2L^2}{p^2}\right)(1 - \rho^p) + \frac{z^2 Var(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L)}{Var(D)}. \tag{2.10}$$

From (2.10) can be observed that the order rate variance ratio becomes higher when a higher z is chosen. However, the exact behaviour of the last term of (2.10) stays unclear, because $Var(\hat{\sigma}_{et}^L - \hat{\sigma}_{e,t-1}^L)$ is unknown and depends on the other parameters. It would be interesting to know a bit more about this term. In the next section we do a spreadsheet simulation with Crystal Ball where we study 'among others' the behaviour of the last term of (2.10).

2.3 Simulation Experiments

In this section we simulate the supply chain of the beer game using Excel with Crystal Ball. First, we simulate the process with exponential smoothing forecasting, a method that is often used in practice. Secondly, we simulate the simple supply chain model of the previous section to study the behaviour of the order rate variance ratio and compare the simulation results with the analytical results derived by Chen et al. [6].

2.3.1 Simulation with Exponential Smoothing Forecasting

Consider a supply chain in which all members observe their inventory at the beginning of each week and place an order to their upstream neighbour based on their observation. We assume that the demands seen by the retailer follow the AR(1) model of (2.1). There is a fixed lead time of L weeks and any unfilled orders are backlogged. We assume further that all members of the supply chain use a simple order-up-to policy, where the order-up-to point y_t is estimated from the observed demand as in Equation (2.4). Suppose that the retailer, wholesaler, manufacturer and supplier use exponential smoothing to forecast the mean demand. The mean lead time demand is in then estimated by

$$\hat{D}_t^L = L(\alpha D_{t-1} + (1-\alpha)\hat{D}_{t-1}). \tag{2.11}$$

To estimate the standard deviation of the forecast error we use the square root of the sample variance of the forecast error based on one year of simulated historical data.
The inventory policy and forecasting technique described above are not necessarily optimal. However, such policies are commonly used in practice and give a good impression of how the bullwhip effect can occur.

Figure 2.3 shows the effects of the policy described above in a typical simulation run. The bullwhip effect is clearly present. In the figure you see that the variance of the orders become greater at each stage further upstream the supply chain. Also the inventory levels contain larger and larger swings at each stage. Consequently, high levels of safety stock are needed. The graphs of Figure 2.3 look different from the graphs presented by Sterman, see Figure 2.2. The effects shown by Sterman look more extreme than our simulation results. This is probably because behavioural causes of the bullwhip effect are not present in our simulation. Members of the supply chain order strictly consistent with an order-up-to policy, where players of the beergame react more directly to demand signals from their upstream neighbours.

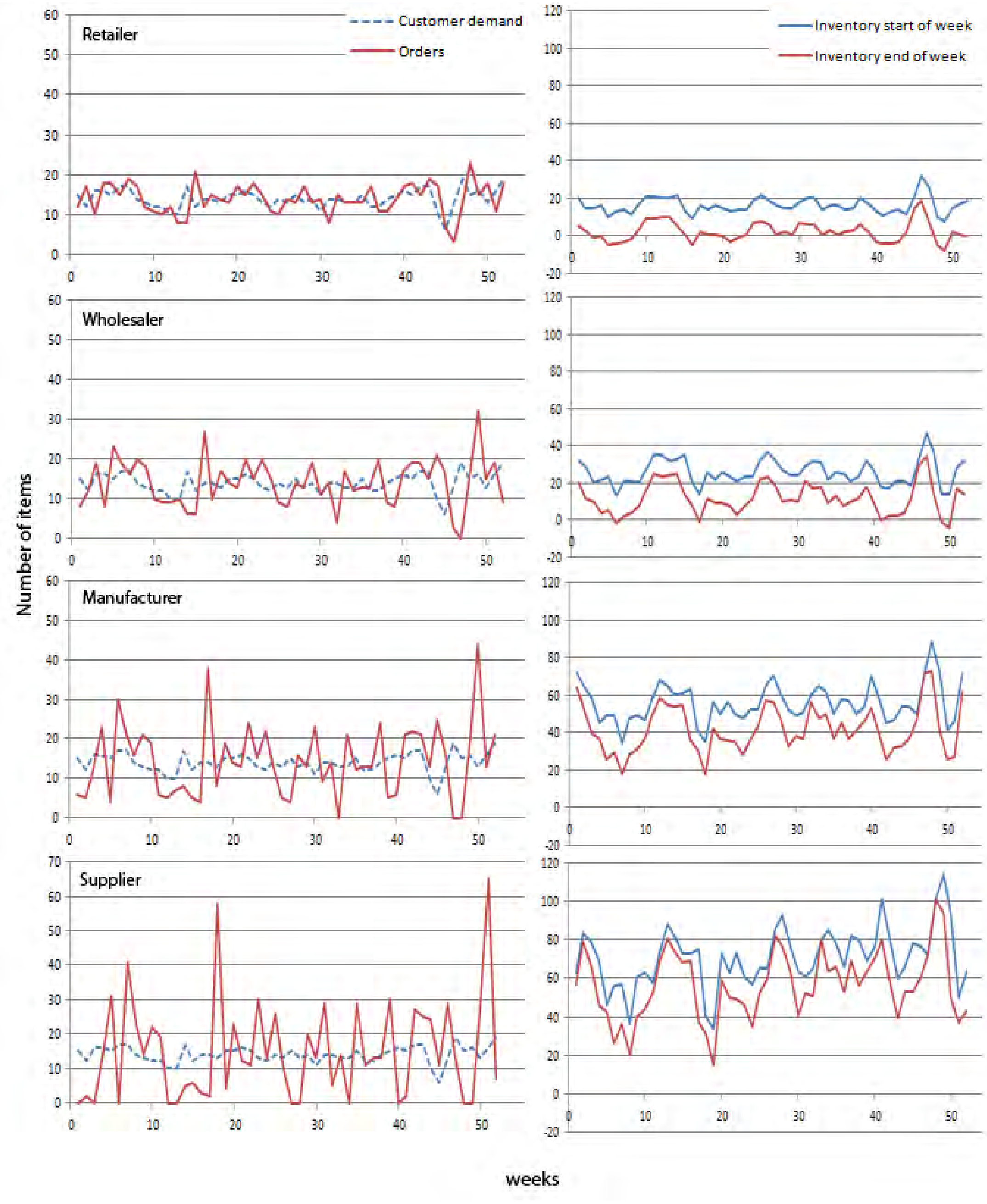


FIGURE 2.3: Simulation of the supply chain model with exponential smoothing. $\mu = 10, \rho = 0.3, \sigma = 2, L = 2$, $z = 2, \alpha = 0.3$

To study the impact of the different parameters of the model, we simulate the model to forecast the order rate variance ratio (ORVR) at the retailer under different parameter values. The order rate variance ratio is given by the variance of the orders divided by the variance of the demand. We only study the situation at the retailer, because the ratio is stable at each stage of the supply chain, so the results will be the same at other links in the supply chain. For each parameter, we vary the value of the parameter, while keeping the other parameters constant. We use the parameter setting of Figure 2.3 as basic setting. Figure 2.4 shows the impact of the different parameters on the order rate variance ratio. The forecasts are based on 1000 simulation runs, where in each simulation run the order rate variance ratio is estimated from 2 year of data.

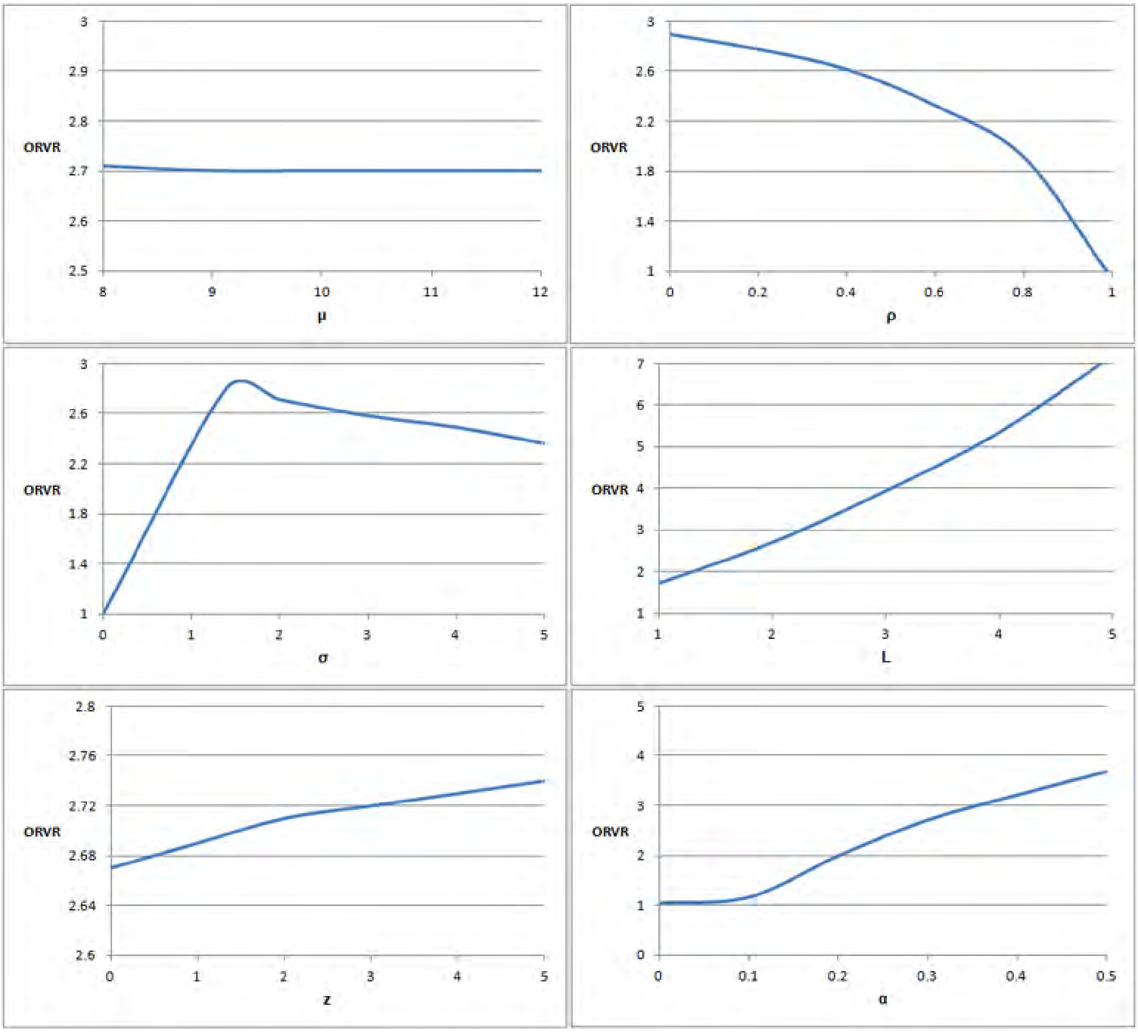


FIGURE 2.4: The impact of different parameters on the order rate variance ratio.

Several observations can be made regarding the impact of the different parameters as shown in Figure 2.4. It must be noted, that the effects of the parameters on the order rate variance ratio are subject to the inventory policy and forecast technique, so they do not have to hold in general.

The impact of μ.
The parameter μ, which is part of the demand model, has very little impact on the order rate variance ratio. Hereby we assume that μ is significantly larger than σ. If this is not the case, the demand distribution becomes skewed to the right, because negative demand is not possible.

The impact of ρ.
For $0 \leq \rho < 1$, the order rate variance ratio is decreasing in ρ, the correlation parameter of the demand model. If ρ increases, the variance of the demand and the variance of the orders increase. However, the variance of demand grows relatively faster than the variance of the orders. Consequently, the order variance ratio drops.

The impact of σ.
If the standard deviation of the demand σ is close to zero, there is no bullwhip effect. In that case the variance of the demand and the variance of the orders are both zero. This is intuitively clear, because when the demand is deterministic, your demand forecasts become perfect. Consequently, the order quantity is every week exactly the same as the demand.
If σ is small, then the variance of the orders grows relatively faster than the variance of the demand. This holds until a certain turnover point depending on the parameter settings. From there on, the order rate variance ratio decreases steadily in σ.

The impact of L.
Figure 2.4 shows the great impact of the lead time parameter L. The variance ratio grows fast if the lead time becomes larger. In that case the retailer has to forecast the demand over a long period, which leads to large forecast errors and a high order variance.

The impact of z.
If the retailer chooses a larger value for the service level z, the variance ratio will grow. However, the effect is small compared to the effect of σ, ρ and L.

The impact of α.
We see in Figure 2.4 that the variance ratio grows fast if we put more weight on the last observed demand. Using the last observed demand makes the demand forecast more volatile and leads in general to a higher order variance. If the demand is stationary, it is better to use the mean and standard deviation of the demand only. However, if the demand for example contains seasonality or a trend, this will lead to high forecast errors and bad customer service. In practical situations, the mean demand is often not stable, but contains seasonality. With strong demand seasonality, it would be good to consider the last observed demand in the demand forecast, e.g use a positive α.

2.3.2 Simulation With Moving Average Forecasting

In this section we simulate the simple supply chain model of Chen et al. [6]., described in Section 2.2. The difference with the previous paragraph is that we now use moving average forecasting, instead of exponential smoothing. Chen et al. use a simple supply chain model to derive a lower bound on the order rate variance ratio (2.9). The lower bound shows that the bullwhip effect exists, but the size of the effect stays unclear. In this simulation experiment we forecast the order rate variance and use it to calculate the fraction of the variance ratio that is contributed by the last term of Equation (2.10). In the experiments we each time vary one parameter while keeping the other parameters constant. As basic setting we use $\mu = 10$, $\rho = 0.3$, $\sigma = 2$, $L = 2$, $z = 2$, $p = 20$. The results are shown in Figure 2.5.

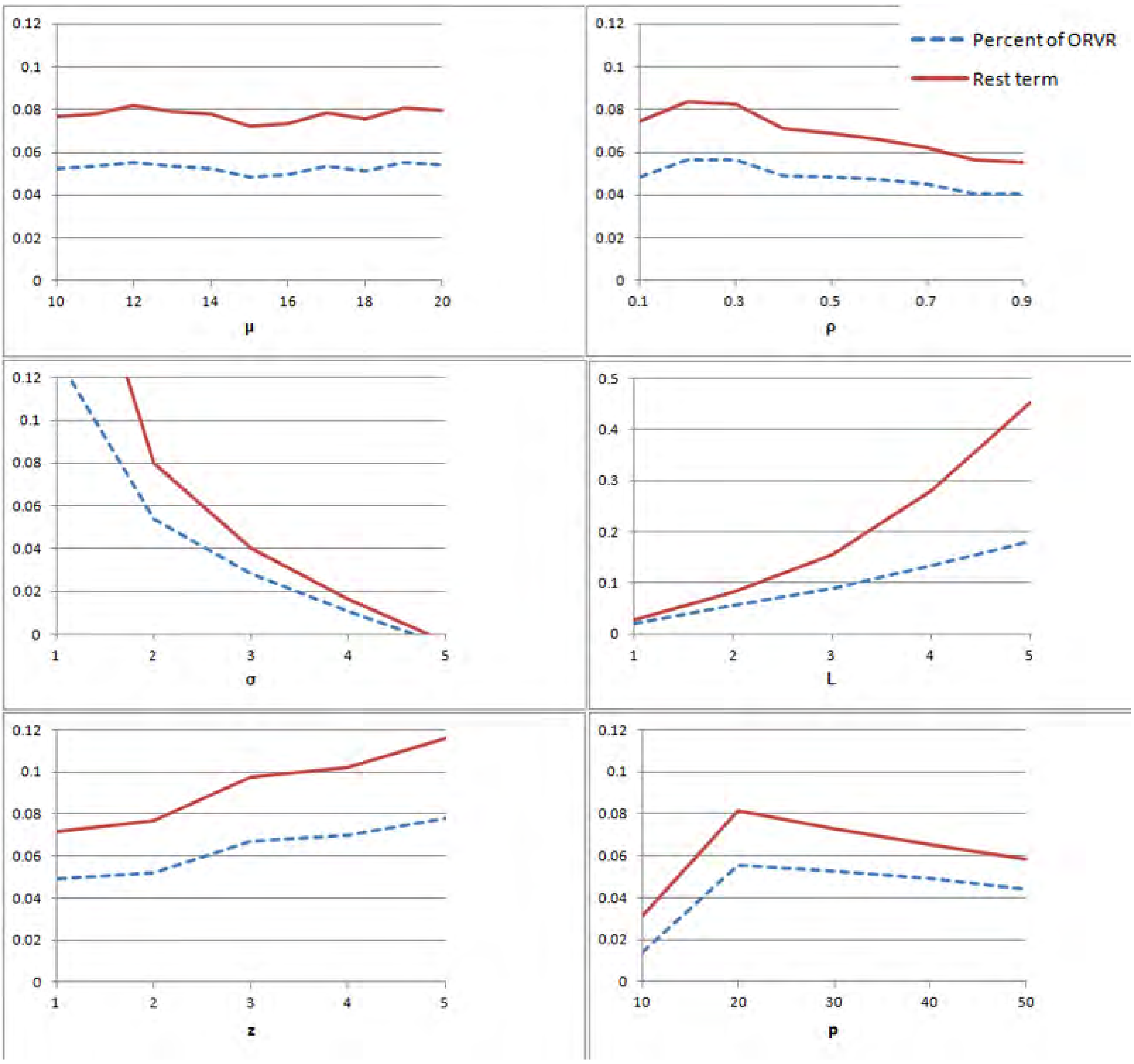


FIGURE 2.5: The impact of different parameters on the last term of Equation (2.10)

The red line in Figure 2.5 shows the absolute value of the last term of Equation (2.10). The blue line is the fraction of this rest term compared to the total size of the order rate variance ratio. We see that in most cases the rest term is smaller than 10% of the variance ratio. However, the rest term starts to grow significantly as σ decreases or if L increases. If σ decreases, the variance of the demand will be smaller and therefore the rest term will be smaller too. If the lead time L becomes larger, then it becomes more difficult to estimate the lead time demand and the mutual differences in forecasts will be larger. Consequently, the numerator of the rest term will grow.

Chapter 3

Causes of the Bullwhip Effect

Originally, the bullwhip effect was ascribed mainly to the irrational behaviour of individuals. Forrester [12] and Sterman [22] both assigned behavioural causes to the bullwhip effect. Later, since Lee et al. [17], much research focused on operational causes of the phenomenon. Lee et al [17] identified four high level causes of the bullwhip effect: demand signal processing, rationing and shortage gaming, order batching and price fluctuation. These causes have been widely accepted as a complete explanation for the bullwhip effect. However, as Sterman [22] showed, behaviour and irrational decision making of individual managers can have a significant impact on the bullwhip effect. This chapter first gives a description of the behavioural causes as suggested by Forrester and Sterman. Secondly, it gives an elaborate description of the four operational causes identified by Lee et al. ([17],[18]) and how these causes can lead to the bullwhip effect.

3.1 Behavioural Causes

Forrester suggested that the decision logic of the individuals that are responsible for demand management creates a tendency to over-respond to increases or decreases in demand from customers in terms of orders placed on the immediate upstream neighbours. If customer demand increases, demand to the upstream neighbour is increased to meet the increase of customer demand. Demand to the upstream neighbour is further increased to replenish inventory, which was depleted by the initial increase of customer demand. The same exaggerated effects can occur if customer demand drops. Next to that, the time lags in the transmission of both order information and materials along the supply chain can exacerbate the effect. Buffa and Miller (1979) illustrated this with the following example. Imagine a product with constant deterministic demand that is delivered through the supply chain of the beergame, depicted in Figure 2.1. The retailer sees a permanent 10% drop in sales on day 1, but consistent with a reorder point policy, does not place an order until say day 10. Accordingly, the wholesaler notes the 10% decrease on day 10, but does not place an order until day 20. As this process moves up the supply chain, the firm furthest upstream may not discover the decline in demand for several weeks. However, during this time they are producing at the old rate, which is $1/0.9 = 111\%$ of the new consumer requirement. Consequently, excess production of 11% per day would have accumulated since day 1. Due to the overstock position, production may be cut back substantially more than 10%, which is an exaggerated reaction to the actual demand decrease.

Sterman [22] also identified a number of behavioural causes of the bullwhip effect. First, members of the supply chain do not adequately account for the delays between order placement and order delivery. These misperceptions of time lags lead to continued over and under ordering in the intervening periods. Moreover, members of the supply chain do not use optimal stock levels and only try to optimise their own element of the chain.

3.2 Operational Causes

Lee et al. ([17], [18]) showed that the bullwhip effect is not solely a result of irrational decision making, but a consequence of the players rational behaviour within the supply chain's infrastructure. Lee et al. [17] identified four high level causes of the bullwhip effect:

1. Demand signal processing
2. Order batching
3. Price fluctuation
4. Rationing and shortage gaming

These four causes have become the standard and are widely accepted as a framework for classifying all causes of the bullwhip effect. In this section we describe these for causes and present the mathematical proofs of Lee et al. [18] that these causes indeed lead to the bullwhip effect.

3.2.1 Demand Signal Processing

The product forecasting in the "beer game setting" is based on the order history from the company's immediate downstream neighbour. Each player in the beer game usually

acts on the demand signals that he or she observes. When a downstream operation places an order, the upstream manager processes that piece of information as a signal about future product demand. Based on this signal, the upstream manager readjust his or her demand forecast. This is what Lee et al. [17] call demand signal processing.
For example, suppose that a manager uses exponential smoothing to determine how much to order from a supplier. Exponential smoothing is commonly used in practice. It is a forecast technique whereby demands are continuously updated as the new daily demand data become available. The forecasts of future demands and associated safety stocks are updated using the smoothing technique. With long lead times, it is not uncommon to have high levels of safety stocks. The fluctuations in order quantities can therefore be much greater than those in demand data. Now, suppose the upstream neighbour in the supply chain also uses exponential smoothing and uses the orders to forecast the demand. The orders that are placed by this neighbour will have even bigger swings, see Figure 3.1.

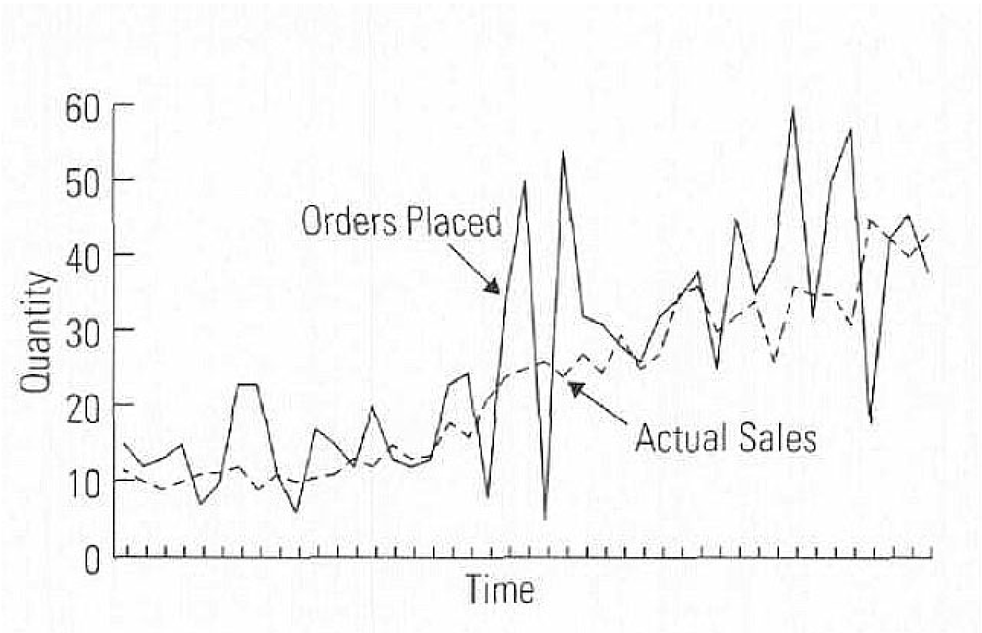


FIGURE 3.1: From Lee et al. [17]. Higher variability in orders from retailer to manufacturer than actual sales.

It is intuitively clear that longer lead times lead to greater fluctuations. As explained above, safety stock contributes to the bullwhip effect. With longer lead times the need for safety stock will be greater.

The reasoning above leads to the hypotheses that the variance of retail orders is strictly larger than that of retail sales and the larger the replenishment lead time, the larger the variance of orders. Lee et al. [18] give a mathematical proof of the hypotheses. A summary is given below.

Mathematical Approach

Consider a multi-period inventory model. Lee et al. [18] assume serially correlated demands which follow the process

$$D_t = d + \rho D_{t-1} + \mu_t \tag{3.1}$$

where D_t is the demand in period t, ρ is a correlation parameter satisfying $0 < \rho < 1$ and μ_t is independent and identically normally distributed with zero mean and variance σ^2. Note that this is a different notation for a similar demand model as used by Chen et al. [6], described in Chapter 1, see Equation (2.1). The only difference is the assumption that μ_t is normally distributed, instead of symmetrically distributed.

The retailer orders a quantity z_t from a supplier each period t. There is a delay of v periods between ordering and receiving the goods, representing the order lead time and the transit time from the supplier to the retailer's site. It is assumed that excess inventory can be returned without cost. Let h, π and c denote the unit holding cost, the unit shortage penalty cost and the unit ordering cost, respectively. S_t is the amount in stock plus on order (including those in transit). Let β be the cost discount factor per period. The cost-minimization problem in an arbitrary period (normalized at 1) is formulated as

$$\min_{S_t} \left[\sum_{t=1}^{\infty} \beta^{t-1} E_1 \left[c z_t + \beta^v g\left(S_t, \sum_{i=t}^{t+v} D_i \right) \right] \right], \tag{3.2}$$

where

$$g\left(S_t, \sum_{i=t}^{t+v} D_i \right) = h\left(S_t - \sum_{i=t}^{t+v} D_i \right)^+ + \pi \left(\sum_{i=t}^{t+v} D_i - S_t \right)^+ .$$

See Heyman et al. [13] for its derivation. The expectation operator E_1 denotes the expectation taken at the decision point of period 1, conditional on the demand realizations $D_i, i = 0, -1, -2, \ldots$.

Theorem 3.1. *In the setting of (3.2), we have:*
(a) if $0 < \rho < 1$, the variance of retail orders is strictly larger than that of retail sales; i.e, $Var(z_1) > Var(D_0)$.
(b) if $0 < \rho < 1$, the larger the replenishment lead time, the larger the variance of orders; i.e., $Var(z_1)$ strictly increases in v.

Proof. (From Lee et al. [18])
Heyman et al. [13] showed that the minimization problem of Equation (3.2) can be solved by solving

$$\min_{S_t} \sum_{t=1}^{\infty} \beta^{t-1} E_1 \left[G(S_t) \right],$$

where $G(S) = c(1-\beta)S + \beta^v E_1 \left[g\left(S, \sum_{i=1}^{v+1} D_i \right) \right]$. Also, they show that

$$S_1^* = Q_{v+1}^{-1} \left[\frac{\pi - c(1-\beta)/\beta^v}{h+\pi} \right], \tag{3.3}$$

where Q_{v+1} denotes the distribution function of $\sum_{i=1}^{v+1} D_i$.
Now note that

$$\begin{aligned} D_k &= d + \rho D_{k-1} + \mu_k \\ &= d(1+\rho) + \rho^2 D_{k-2} + (\rho\mu_{k-1} + \mu_k) \\ &= \ldots \\ &= d\frac{1-\rho^k}{1-\rho} + \rho^k D_0 + \sum_{i=1}^{k} \rho^{k-i}\mu_i \qquad \text{for } k \geq 1. \end{aligned}$$

Thus $\sum_{i=1}^{v+1} D_i$ at the decision point in period 1 is a $N(M, \Sigma)$ distributed random variable with

$$M := d \sum_{k=1}^{v+1} \frac{1-\rho^k}{1-\rho} + \frac{\rho(1-\rho^{v+1})}{1-\rho} D_0,$$

and

$$\Sigma := \sum_{k=1}^{v+1} \sum_{i=1}^{k} \rho^{2(k-i)} \sigma^2.$$

Thus, from (3.3), we have:

$$S_1^* = d \sum_{k=1}^{v+1} \frac{1-\rho^k}{1-\rho} + \frac{\rho(1-\rho^{v+1})}{1-\rho} D_0 + K^* \sigma \sqrt{\sum_{k=1}^{v+1} \sum_{i=1}^{k} \rho^{2(k-i)}}, \tag{3.4}$$

where

$$K^* = \Phi^{-1} \left(\frac{\pi - c(1-\beta)/\beta^v}{h+\pi} \right),$$

for the standard normal distribution function Φ. From (3.4), the optimal order amount z_1^* is given by

$$\begin{aligned} z_1^* &= S_1^* - S_0^* + D_0 \\ &= \frac{\rho(1-\rho^{v+1})}{1-\rho}(D_0 - D_{-1}) + D_0. \end{aligned} \tag{3.5}$$

It follows that

$$Var(z_1^*) = Var(D_0) + \left(\frac{\rho(1-\rho^{v+1})}{1-\rho}\right)^2 Var(D_0 - D_{-1}) + 2\frac{\rho(1-\rho^{v+1})}{1-\rho} Cov(D_0 - D_{-1}, D_0). \quad (3.6)$$

Noting the independence between D_{-1} and μ_0, it can be shown that $Var(D_0) = Var(D_1) = \sigma^2/(1-\rho^2)$, $Var(D_0 - D_{-1}) = 2\sigma^2/(1+\rho)$, and $Cov(D_0 - D_{-1}, D_0) = \sigma^2/(1+\rho)$. Hence,

$$Var(z_1) = Var(D_0) + \frac{2\rho(1-\rho^{v+1})(1-\rho^{v+2})}{(1+\rho)(1-\rho)^2}\sigma^2 > Var(D_0). \quad (3.7)$$

This proves part (a) of Theorem 3.1. Part (b) is also straightforward from (3.7). □

The theorem states that the variance amplification takes place when the retailer adjusts the order-up-to level based on the demand signals. Also, the degree of amplification increases in the replenishment lead time. Moreover, from (3.7) can be observed that when $0 < \rho < 1$ and $v = 0$, we have $Var(z_1) = Var(D_0) + 2\rho\sigma^2$, showing that the amplification effect exists, even when the lead time is zero.

The order rate variance ratio is given by

$$\frac{Var(z_1)}{Var(D_0)} = 1 + \frac{2\rho(1-\rho^{v+1})(1-\rho^{v+2})}{Var(D_0)(1+\rho)(1-\rho)^2}\sigma^2. \quad (3.8)$$

Note that in the simple supply chain model of Chapter 1 we also saw that the amplification effect always exists and that the effect becomes stronger when the lead time increases. However, the size of the effect differs due to the different methods that were used. The order rate variance ratio in Chapter 1 (2.10), was based on a commonly used method to determine the order up to point (2.4), and a simple moving average to estimate the mean and variance of demand. These methods are not optimal. In contrast, the order rate variance ratio in this section is a result of the optimal inventory policy given the ordering costs, unit shortage costs and unit holding costs.

3.2.2 Order Batching

In a supply chain, most companies batches or accumulate demands before issuing an order. Instead of ordering frequently, companies may order weekly, biweekly or even monthly. There are many reasons for ordering in batches. For example, a company might order a full truck or container load from its supplier to receive a quantity discount and minimize transport costs. Many manufacturers order from their suppliers after they ran their material requirements planning (MRP) systems. These (MRP) systems often run once a month, resulting in a highly erratic stream of orders, see Figure 3.2. When a company faces periodic ordering by its downstream neighbour, it sees a higher variability in demand than the downstream neighbour itself. Periodic ordering amplifies variability and will therefore contribute to the bullwhip effect. This effect is small if all customers' order cycles were spread out evenly throughout time in a deterministic way. Unfortunately, orders are more likely to be randomly spread out or worse, to be correlated. When order cycles are correlated, most customers order at more or less the same time. This results in even higher peaks and higher variability.

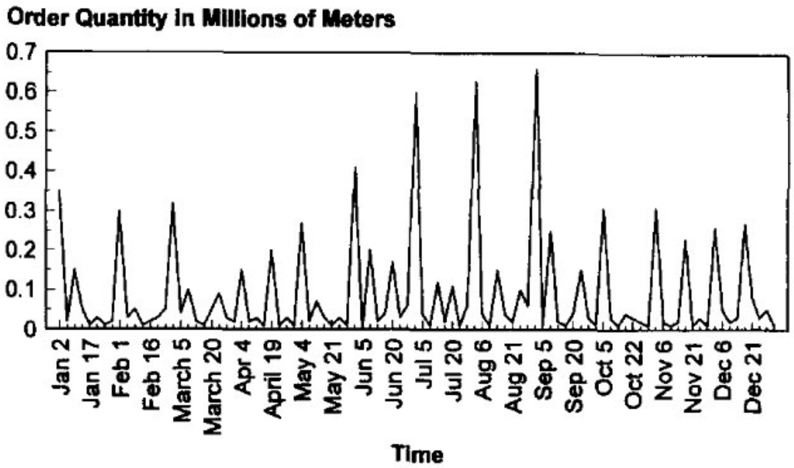


FIGURE 3.2: From Metters [19]. Orders from auto-mobile parts manufacturers to their supplier, a tubing manufacturer. The spikes coincidence with the MRP runs of the manufacturers.

Mathematical Approach

Lee et al. [18] considers a periodic review system with stationary demand and full backlogging at a retailer. Suppose there exist N retailers each using a periodic review system with the review cycle equal to R periods. Suppose that the demands for retailer j in period k is ξ_{jk}, which is i.i.d with mean m and variance σ^2 for each retailer. Lee [18] considers three cases.

Case 1. Random Ordering. Demands from retailers are independent.

Case 2. Positively Correlated Ordering. The extreme case is considered in which all retailers order in the same period.

Case 3. Balanced Ordering. The extreme case in which orders from retailers are evenly distributed in time.

Case 1. Random ordering. Let n be a random variable denoting the number of retailers that order in a randomly chosen period. This means that n is a binomial random variable with parameters N and $1/R$. Hence, $E(n) = N/R$ and $Var(n) = N(1/R)(1-1/R)$. Let Z_t^r denote the total order quantity from n retailers in period t, i.e.,

$$Z_t^r := \sum_{j=1}^{n} \sum_{k=t-R}^{t-1} \xi_{jk}. \tag{3.9}$$

Then we have

$$E(Z_t^r) \quad = \quad E[E(Z_t^r|n)] = E[nRm] = Nm, \tag{3.10}$$

and

$$\begin{aligned} Var(Z_t^r) &= E[Var(Z_t^r|n)] + Var[E(Z_t^r|n)] \\ &= N\sigma^2 + R^2m^2\frac{N}{R}\left(1-\frac{1}{R}\right) \\ &= N\sigma^2 + m^2N(R-1) \geq N\sigma^2 \end{aligned} \tag{3.11}$$

If $R = 1$, the variance of the demand seen by the supplier would be the same as the demand seen by the retailer. If R increases, the variance of the demand seen by the supplier increases. This shows that in this case, order batching leads to the bullwhip effect. The variance also increases if the number of retailers N increase.

Case 2. Positively Correlated Ordering. Al retailers order in the same period of the review cycle with probability $1/R$, so

$$Pr(n=i) = \begin{cases} 1-1/R & \text{for } i=0 \\ 1/R, & \text{for } i=N \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

Here, n is a random variable with

$$E(n) = N/R$$

and

$$Var(n) = \frac{N^2}{R(1-1/R)}.$$

Now define Z_t^c as in (3.9), then

$$E(Z_t^c) \quad = \quad E[E(Z_t^c|n)] = E[nRm] = Nm, \tag{3.13}$$

and

$$\begin{aligned} Var(Z_t^c) &= E[Var(Z_t^c|n)] + Var[E(Z_t^c|n)] \\ &= N\sigma^2 + R^2m^2\frac{N^2}{R}\left(1-\frac{1}{R}\right) \\ &= N\sigma^2 + m^2N^2(R-1) \geq N\sigma^2. \end{aligned} \tag{3.14}$$

The extra variance of the total order quantity Z_t^c is N times greater than it was in the case of random ordering, see Equation (3.11). This means that the number of retailers has even a larger impact on the bullwhip effect than it has in the case of random ordering.

Case 3. Balanced Ordering. To analyse this case, Lee et al. [18] introduce the following scheme to evenly distribute the orders in time. Suppose $N = MR + k$, where M and k are integers, N denotes the total number of retailers and R denotes the number of periods in the retailers' review cycle with $0 \leq k \leq R$. The retailers are divided into R groups: k groups of size $M+1$ and $R-k$ groups each of size M. Then, all retailers in the same group place orders in a designated period within a review cycle, and no two groups order in the same period. For example, if $R = 7$ and $N = 33$ then $33 = M*7+k$ with $0 \leq k \leq 7$. So $(N, M, R, k) = (33, 4, 7, 5)$. Thus, 5 groups of 5 retailers and 2 groups of 4 retailers. Group $1, \ldots, 7$ orders at day $1, \ldots, R$, respectively. Each retailer places an order with chance $1/R$.
Here,

$$Pr(n = i) = \begin{cases} 1 - k/R & \text{for } i = M \\ k/R, & \text{for } i = M+1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

The mean and variance are in this case given by

$$E(n) \quad = \quad M\left(1-\frac{k}{R}\right) + (M+1)\frac{k}{R} = \frac{N}{R}, \tag{3.16}$$

and

$$\begin{aligned} Var(n) &= \left(1-\frac{k}{R}\right)M^2 + (M+1)^2\frac{k}{R} - \left(\frac{N}{R}\right)^2 \\ &= \left(\frac{k}{R}\right)\left(1-\frac{k}{R}\right). \end{aligned} \tag{3.17}$$

Now define Z_t^b as in (3.9), then

$$E(Z_t^b) = E[E(Z_t^b|n)] = E[nRm] = Nm, \tag{3.18}$$

and

$$\begin{aligned} Var(Z_t^b) &= N\sigma^2 + R^2m^2\frac{k}{R}\left(1 - \frac{k}{R}\right) \\ &= N\sigma^2 + m^2k(R-k) \end{aligned} \tag{3.19}$$

Note that if the number of retailers N can be balanced completely, then k is zero and the last term of (3.19) vanishes. Moreover, the order rate variance ratio in the case of balanced ordering is

$$\frac{Var(Z_t^b)}{Var(D_t)} = 1 + \frac{m^2k(R-k)}{N\sigma^2}, \tag{3.20}$$

which means that the increase in variance decreases in N. This is because more retailers lead to a better balance.

In any case, since $k(R-k) \leq N(R-1)$ for each $k = 1, 2, \ldots, R$,

$$N\sigma^2 + m^2k(R-k) \leq N\sigma^2 + m^2N(R-1) \leq N\sigma^2 + m^2N^2(R-1),$$

which leads to the following theorem.

Theorem 3.2. *Let Z_t^c, Z_t^r, and Z_t^b be random variables denoting the orders from N retailers under correlated ordering, random ordering and balanced ordering respectively. Then,*
(a) $E[Z_t^c] = E[Z_t^r] = E[Z_t^b] = Nm$,
(b) $Var(Z_t^c) \geq Var(Z_t^r) \geq Var(Z_t^b) \geq N\sigma^2$.

The theorem confirms that the variability of demand as seen by the supplier is higher than that experienced by the retailers. Thus, order batching leads to the bullwhip effect. The effect is the strongest when the periods in which retailers place their orders are correlated. The effect is weaker, when these periods are random and the weakest if they are evenly distibuted in time. If the order periods can be balanced completely, i.e. if N is a multiple of R, the bullwhip effect due to order batching does not occur.

3.2.3 Price Fluctuation

Companies often buy items in advance of requirements. This "forward buying" results from price fluctuations due to special promotions like price discounts, quantity discounts and coupons. The result is that companies buy in quantities that do not reflect their immediate needs. They often buy in bigger quantities and stock up for the future. If the cost of holding inventory is less than the price difference, buying in advance may well be a rational decision. However, when companies buy more than needed and wait until their inventory is depleted, the variation of the buying quantities is much bigger than the variation of the consumption rate, which leads to the bullwhip effect.

An example of sale and order patterns of chicken noodle soup from Lee et al. [17] shows how high-low buying practices can lead to high variability in shipments from manufacturer to distributors, see Figure 3.3. Such wide swings often force companies to run their factories overtime at certain times and to be idle at others. Alternatively, companies may have to build huge piles of inventory to anticipate on big swings in demand. On the other hand, if the manufacturer would not do price discounting, a competitor who does, might take over its business.

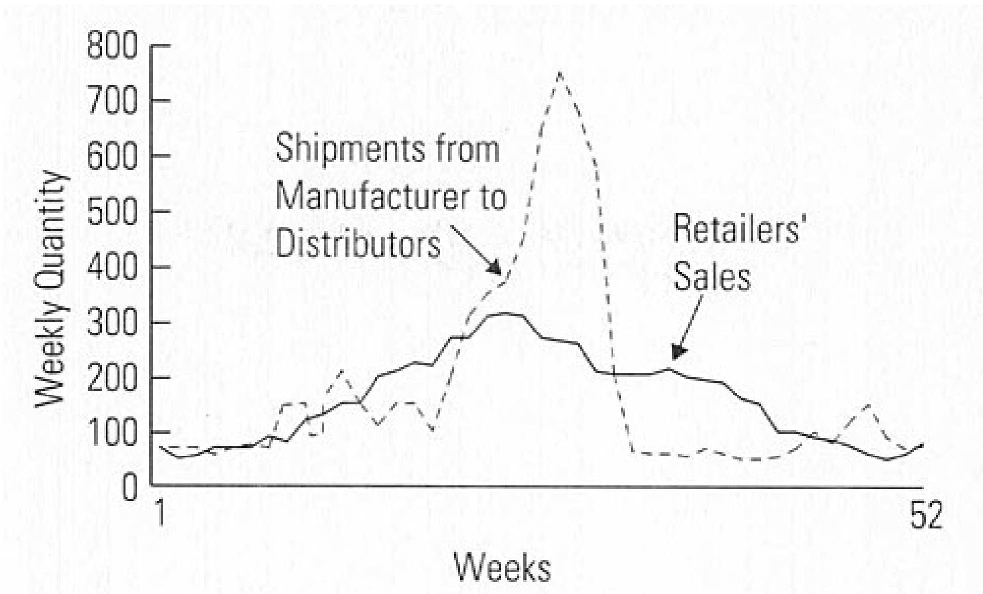


FIGURE 3.3: From Lee et al. [17]. Bullwhip effect in the supply chain of chicken noodle soup.

Mathematical Approach

Lee et al. [18] analyze price fluctuations as follows. Suppose the retailer faces each period independent and identically distributed demand with density function $\phi(\cdot)$. The purchase price for its product alternates over time between c^L and c^H, where $c^L < c^H$. With probability q the price in a period will be c^L and with probability $1 - q$ the price will be c^H. In this situation, the retailer's inventory problem can be formulated as

$$V^i(x) = \min_{y \geq x} \left[c^i(y-x) + L(y) + \beta \int_0^\infty [qV^L(y-\xi) + (1-q)V^H(y-\xi)]\phi(\xi)d\xi \right],$$

where

$$L(y) = p \int_y^\infty (\xi - y)\phi(\xi)d\xi + h \int_0^y (y-\xi)\phi(\xi)d\xi. \tag{3.21}$$

Here V^i denotes the minimal expected discounted cost incurred throughout an infinite horizon when the current price is c^i, $i \epsilon L, H$, $L(\cdot)$ is the sum of one-period inventory and shortage costs at a given level of inventory. The unit shortage penalty cost is denoted by p and the unit holding cost by h.
Without giving a proof of the result, Lee et al. [18] report the following optimal inventory policy to problem (3.21).

At price c^L, get as close as possible to the stock level S^L. At price C^H, bring the stock level to S^H, where $S^H < S^L$.

To show the causes of price fluctuations, we use the following example. Suppose that the cost parameters are chosen such that $S^H = 30$ and $S^L = 100$. In period 0 the manufacturer's price is low, and the retailer raises its inventory level to 100. For a number of periods, the manufacturer's price remains high, so the retailer does not order until the inventory level hits below 30. The retailer orders up to 30 until the manufacturer drops its price to c^L. When this happens, the retailer orders up to 100 items until the price is raised. Here we simulated this example with normally distributed demand with $\mu = 15$ and $\sigma = 5$. The chance q on a low purchase price in a certain period is 10%. Figure 3.4 shows a typical result of a simulation run. The high peaks represent periods with a low purchase price. The bullwhip effect can be clearly observed.

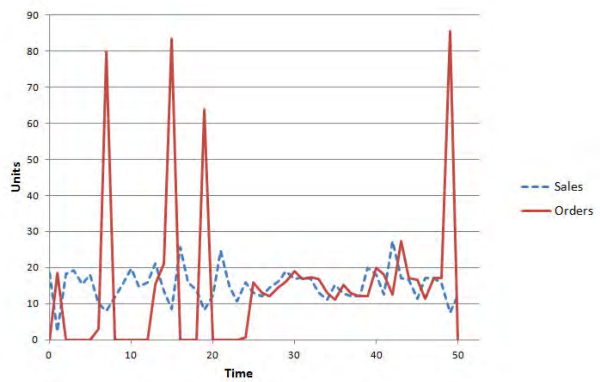


Figure 3.4: Bullwhip effect due to price fluctations and rational decision making.

3.2.4 Rationing and Shortage Gaming

When product demand exceeds supply, a manufacturer often rations its product to customers. For example, the manufacturer then allocates its products in proportion to the amount ordered by the different retailers. Retailers often anticipate on potential shortages by exaggerating their real needs when they order. If demand drops later, this will lead to small orders and cancellations. Lee et al. [17] call this overreaction by customers, rationing and shortage gaming. This "gaming" results in misleading information on the product's real demand. To illustrate the effects of rationing gaming on the variance amplification, consider a supply chain consisting of a manufacturer, multiple wholesalers, and multiple retailers. If the manufacturer appears to be in short of supply, wholesalers will play the rationing game to get a large share of the supply. Assessing a possibility of the wholesaler not getting enough from the manufacturer, retailers also play the rationing game. The effect is that demand and its variance are amplified as one moves up the supply chain. In practice, there are many examples of this rationing and shortage gaming. One example is the shortage of DRAM chips in the 1980's, from Lee et al. [18]. In the computer industry, orders for these chips grew fast, not because a growth in customer demand, but because of anticipation. Customers placed duplicate orders with multiple suppliers and bought from the first one that could deliver, then cancelled all other duplicate orders.

Mathematical Approach

Lee et al. [18] developed a one-period model with multiple retailers to illustrate that retailers rationally will issue an order which exceeds in quantity what the retailer would order if the supply of the product is unlimited. A description is given below.

A manufacturer supplies a single product to N identical retailers indexed by $n = 1, 2, \ldots, N$. Retailer n first observes its demand distribution $\Phi(\cdot)$ and places an order z_n at time 0. Then, the manufacturer delivers the product at time 1. The manufacturer's output μ is a random variable, distributed according to $F(\cdot)$.

In case the total amount of orders $\sum_{n=1}^{N} z_n$ exceeds the realized output μ, the manufacturer allocates the output to retailers in proportion to their orders. The total retail orders are denoted by $Q := \sum_j z_j$. If the realized capacity μ is smaller than Q, retailer i will receive $z_i\mu/Q$ due to the allocation.

Let $C_i(z_1, \ldots, z_i, \ldots, z_N)$ denote the expected cost for retailer i when retailer i chooses the order quantity z_i for $i = 1, 2, \ldots, N$. A Nash equilibrium is defined as the order quantities $(z_1^*, z_2^*, \ldots, z_N^*)$ chosen by retailers who each take the decisions of others as given and choose z_i to minimize the expected cost. That is, $(z_1^*, z_2^*, \ldots, z_N^*)$ is a Nash equilibrium if, for each $i = 1, 2, \ldots, N$,

$$z_i^* = \arg\min_{z_i} C_i(z_1, \ldots, z_{i-1}, z_i, z_{i+1} \ldots, z_N). \tag{3.22}$$

Thus, if the first-order condition approach is valid, the equilibrium must satisfy, for each i,

$$\frac{dC_i}{dz_i}(z_1, \ldots, z_{i-1}, z_i, z_{i+1} \ldots, z_N)|_{z_i = z_i^*} = 0. \tag{3.23}$$

Since every retailer is symmetric, we focus on the symmetric Nash equilibrium where $z_i^* = z^*$ for each $i = 1, 2, \ldots, N$. That is, z^* satisfies

$$\frac{dC_i}{dz_i}(z^*, \ldots, z^*, z_i, z^* \ldots, z^*)|_{z_i = z_i^*} = 0. \tag{3.24}$$

To derive the equilibrium, consider retailer i who takes the other retailers' ordering strategies z_j^* $(j \neq i)$ as given and chooses z_i to minimize the expected cost

$$\begin{aligned} C_i \quad = \quad & \int_{\mu=0}^{Q} \left[p \int_{\mu z_i/Q}^{\infty} \left(\xi - \frac{\mu z_i}{Q} \right) d\Phi(\xi) + h \int_{0}^{\mu z_i/Q} \left(\frac{\mu z_i}{Q} - \xi \right) d\Phi(\xi) \right] dF(\mu) \\ & + (1 - F(Q)) \left[p \int_{z_i}^{\infty} (\xi - z_i) d\Phi(\xi) + h \int_{0}^{z_i} (z_i - \xi) d\Phi(\xi) \right], \end{aligned} \tag{3.25}$$

where $Q = \sum_{j \neq i} z_j^* + z_i$ and the unit shortage penalty cost is p. Note that the decision z_i must be taken before the capacity μ is realized. Hence, there are two possible scenarios depending on the supply condition. One is when the supply μ falls short of the total demand Q, and retailer i is allocated the amount $z_i \mu / Q$. The first term on the RHS of (3.25) is the expected cost conditioned on the manufacturer's output μ in the shortage scenario. The other scenario is when the realized supply is sufficient to meet the total demand. The second term on the RHS of (3.25) is the expected cost times the probability of this scenario.

Its first order condition is given by

$$\begin{aligned} \frac{dC_i}{dz_i} \quad = \quad & \int_{\mu=0}^{Q} \left[-p + (p+h)\Phi\left(\frac{\mu z_i}{Q} \right) \right] \mu \left(\frac{1}{Q} - \frac{z_i}{Q^2} \right) dF(\mu) \\ & + [1 - F(Q)][-p + (p+h)\Phi(z_i)] = 0, \end{aligned} \tag{3.26}$$

where we used Leibnitz's rule and $dQ/dz_i = 1$.

A pseudo-convex function is a function that behaves like a convex function with respect to finding its local minima. If C_i is pseudo-convex, then the solution z_i^0 to Equation (3.26), is the optimal order quantity z_i^*. To establish the pseudo-convexity of C_i, consider

C_i at z_i^0 satisfying Equation (3.24). It must be that $-p + (p+h)\Phi(z_i^0) \geq 0$, because otherwise, we would have that $-p + (p+h)\Phi(\mu z_i^0/Q) \leq 0$ for all $\mu \leq Q$, resulting in $dC_i/dz_i < 0$. Now, consider the second derivative of C_i w.r.t. z_i^0:

$$\frac{d^2C_i}{dz_i^2} = [-p + (p+h)\Phi(z_i^0)]f(Q) + [1 - F(Q)](p+h)\phi(z_i^0) \geq 0. \tag{3.27}$$

This establishes the pseudo-convexity of C_i. It follows that the solution z_i^0 to Equation (3.26) is the optimal order quantity z_i^*.

From (3.24) and (3.26), therefore, the symmetric equilibrium z^* must satisfy

$$\int_0^{N\cdot z_*} \left[-p + (p+h)\Phi\left(\frac{\mu}{N}\right)\right]\mu\left(\frac{1}{N\cdot z^*} - \frac{1}{N^2\cdot z^*}\right)dF(\mu) \\ +[1 - F(z^*\cdot N)][-p + (p+h)\Phi(z^*)] = 0. \tag{3.28}$$

Theorem 3.3. *In the above setting, $z' \leq z^*$, where z' is the solution of the news-vendor problem: $-p + (p+h)\Phi(z)$. The news-vendor problem is the problem with unlimited supply. Further, if $F(\cdot)$ and $\Phi(\cdot)$ are strictly increasing, the inequality strictly holds.*

The theorem states that the optimal order quantity for the retailer in the rationing game exceeds the order quantity in the traditional news-vendor problem. This implies the bullwhip effect when the mean demand changes over time.

Chapter 4

Remedies to the Bullwhip Effect

Lee et al. [17][18] described four operational causes of the bullwhip effect. In Chapter 3 we gave an elaborate description of these causes. This chapter is concerned with proposed remedies to counteract these four causes.

4.1 Remedies to Demand Signal Processing

Repetitive processing of demand data happens when supply chain members process the demand signals from their immediate downstream neighbours. In this case, supply chain members use the demand forecasts of their downstream neighbour to do their own demand forecasts. In Paragraph 3.2.1 we showed how this could lead to the bullwhip effect. In this paragraph, we present three commonly proposed remedies to the problem: collaboration, replenishment smoothing, and operational efficiency.

4.1.1 Collaboration

Probably the most obvious remedy to the problem is collaboration between the supply chain members. Canella et al. [4] define collaboration in a supply chain as "transforming suboptimal solutions of individual links into a comprehensive solution through sharing customer and operational information". Several authors have shown how collaboration could reduce the amplification of orders in the upstream direction (Chen et al. [6], Disney et al. [9], Chatfield et al. [5]), reduce inventory holding costs (Shang et al. [21], Kelepouris et al. [16]), and improve customer service levels (Hosada et al. [15]).

A framework for different supply chain structures based on the degree of collaboration was provided by Holweg et al. [14]. In their model they use the collaboration on inventory replenishment and the collaboration on forecasting as dimensions, see Figure 4.1.

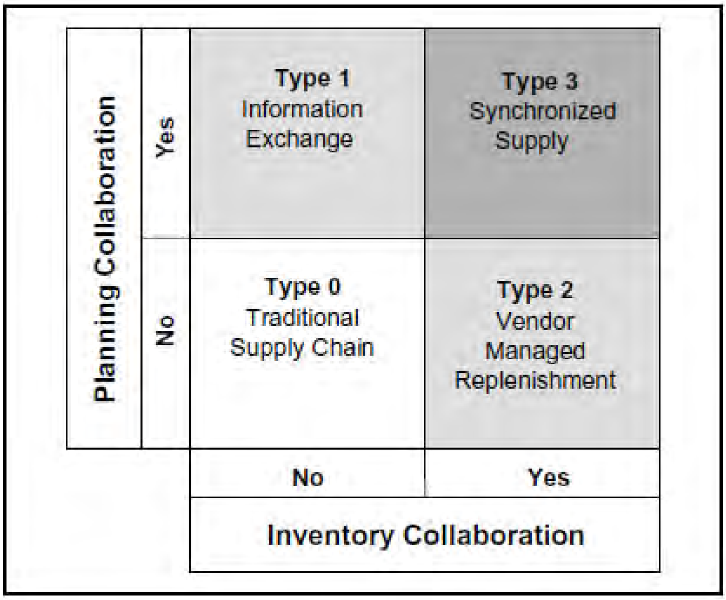


FIGURE 4.1: From Holweg et al. [14]. Supply chain collaboration framework.

The Traditional Supply Chain means that each level in the supply chain issues production orders and replenishes stock without considering the situation at either up- or downstream tiers of the supply chain.

In traditional supply chains the only information that is available to the upstream member are the purchase orders issued by the downstream member. This is the setting of Sterman's beer game [22], described in Chapter 2. In Chapter 3 we saw that in this setting the bullwhip effect will occur, even with rational decision making.

Information Exchange means that individual supply chain members still order independently, yet exchange demand information and action plans in order to align their forecasts for capacity and long-term planning.

Taking end customer sales into consideration at levels further upstream the supply chain is a major improvement over simply relying on the orders sent by the downstream neighbour. Delays in receiving the demand signals are removed, and unnecessary uncertainty is eliminated.

Vendor Managed Replenishment means that the task of generating the replenishment order is given to the supplier, who then takes responsibility for maintaining the retailer's inventory, and subsequently, the retailer's service levels.

Vendor Managed Replenishment (VMR), also often referred to as Vendor Managed Inventory (VMI) offers two sources of bullwhip reduction. One layer of decision making is eliminated and the delays in information flow are reduced.

Synchronized Supply merges the replenishment decision with the production and materials planning of the supplier. Here, the supplier takes charge of the customer's inventory replenishment on the operational level, and uses this visibility in planning his own supply operations.

Collaboration on inventory replenishment and forecasting gives suppliers a better understanding and ability to cope with demand variability. This is an important feature when trying to counter the bullwhip effect.

4.1.2 Replenishment smoothing

In practical situations, a commonly used periodic review policy is the (S, R) order policy. At each review period R the inventory is reviewed and a quantity O is ordered to bring the level of the available inventory to the order-up-to level S. The available inventory consists of the inventory on hand plus the inventory on order but not yet arrived (Work In Progress or Pipeline Inventory).
A smoothing replenishment rule is an (S, R) policy in which the entire gap between the level S and the available inventory is not recovered in one review period. Instead, for each review period R the quantity O is ordered to recover only a fraction of the gap between the target on-hand inventory and the current level of on-hand inventory, and a fraction of the gap between the target pipeline inventory and the current level of pipeline inventory. The fractions to recover are regulated by decision parameters known as proportional controllers. The proportional controller Ty modulates the recovery of the on-hand inventory gap. The proportional controller Tw determines the recovery of the work in progress inventory gap.
Smoothing replenishment is an effective remedy to demand signal processing, because it limits the over-reaction/under-reaction to changes in demand. In the literature it is shown that properly tuning the value of the proportional controllers can reduce the bullwhip effect (Disney and Towill [10], Disney et al. [11]). However, smoothing replenishment rules may have a negative impact on customer service (Dejonckheere et al. [7]). This has only been shown in supply chains with no collaboration.

4.1.3 Operational Efficiency

Many authors showed that long resupply lead times have a bad influence on the bullwhip effect (Lee et al. [18], Chen et al. [6]). These lead times can be reduced by improving the operational efficiency along the supply chain. An effective way to achieve this is with *just-in-time management* (JIT). JIT is a method for inventory control that is part of the *lean manufacturing system*. The philosophy of JIT is that the storage of unused inventory is a waste of resources. To make the elimination of these inventory buffers possible, the focus is on having the right material, at the right time, at the right place, and in the exact amount. JIT controls production in a totally different way than MRP does. MRP systems often run once a month and lead to large monthly orders. With JIT, inventory spaces are reduced, which leads to frequent small orders and thus to short lead times and smaller order batches. A disadvantage of JIT is that a small disruption, for example a small change in demand or a broken machine, can immediately lead to problems because the safety stocks are removed. On the other hand it makes quick reactions possible because small disruptions are quickly observed.

4.2 Remedies to Order Batching

The second cause described by Lee et al. [17] is order batching. As shown by Lee et al. [18], large order batches and low order frequencies contribute to the bullwhip effect, and the effect is amplified if the order periods of different retailers are correlated.

In the previous section we described how the JIT method can counteract large order batches and low order frequencies. However, small order batches and low order frequencies are expensive. One reason is that the relative cost of placing an order and replenishing it becomes higher if the order size decreases and/or if the order frequency increases. A remedy is the use of Electronic Data Interchange (EDI), which is a method for transferring data between different computer systems or computer networks. EDI can reduce the cost of the paperwork in generating an order, which leads to more frequent ordering by customers.
Another reason for large order batches is the otherwise high cost of transportation. Full truckloads are often much more economical than less-than-full truckloads. A remedy is the use of a third party logistics provider (3PL). A 3PL is a is firm that provides multiple logistics services, under which transportation services that can be scaled and customized to the needs of the customer. Third party logistics providers have many customers and can combine orders of different customers, which makes it possible to order less-than-full truckloads for an economical price.

To reduce the amplification effect of correlated ordering, manufacturers could coordinate their resupply with their customers. A manufacturer could for example spread the regular delivery appointments with its customers as much as possible evenly over the week.

4.3 Remedies to Price Fluctation

Companies sometimes buy items in advance of requirements in reaction to price fluctuations. These price fluctuations often result from promotions like price discounts, quantity discounts and coupons. Lee et al. [17] found that the costs of such practices often outweigh the benefits and therefore it is better to stabilize prices. A common way to do this is with a *every day low price policy* (EDLP). With EDLP companies offer no discounts but promise a stable low price which saves customers the effort and expense connected to price fluctuations. In practice, many companies use price discounts to compete with other companies. Therefore, EDLP is for many companies not an option.

4.4 Remedies to Rationing and Shortage Gaming

When product demand exceeds supply, a manufacturer rations its product to customers. In this situation the manufacturer often allocates its products in proportion to the amount ordered by the different retailers. Retailers often anticipate on potential shortages by exaggerating their real needs when they order. A remedy proposed by Lee et al. [17] is to allocate in proportion to past sales records, instead of allocating in proportion to the amounts ordered. Customers then have no incentive to exaggerate their orders.

Another problem is that if demand drops later, companies can easily cancel their orders due to the generous return policy of manufacturers. Without a penalty, retailers will continue to exaggerate their needs and cancel their orders in case they ordered too much. Lee et al. [17] therefore propose that manufacturers enforce more stringent cancellation policies, so that retailers are more cautious with exaggerating their orders.

Chapter 5

Conclusion and Discussion

The bullwhip effect is an obstinate phenomenon that will always be present in supply chains and cannot be completely eliminated. Since Forrester's pioneering work [12] in 1961, much research has been done on the bullwhip effect. Many authors across a range of academic disciplines showed the existence of the effect. Some authors ascribed solely behavioural causes to the bullwhip effect. These causes are the misperception of time legs, which lead to continued over- and underordering, and the use of suboptimal inventory policies and forecasting techniques. Since two important papers of Lee et al. [17] [18] in 1997, four operational causes are widely accepted as a complete explanation for the bullwhip effect. These causes are demand signal processing, rationing and shortage gaming, order batching and price fluctuation. Also, much research has been done on the impact of different inventory policies, forecasting techniques, and other factors that influence the process. These factors include the impact of forecast errors, lead times, collaboration, seasonality, etcetera. The research on the bullwhip effect has led 'among others' to a good explanation of the causes of the bullwhip effect and consequently to many possible remedies.

Since the papers of Lee et al. [17] [18], there is little discussion about the causes of the bullwhip effect. Unfortunately, countering the effect is in many practical situations not that simple. Despite of all research that has been done on the subject, it remains difficult to forecast the exact impact of certain changes or decisions in general. There are many factors that have impact on the bullwhip effect and these factors have often contradictory impacts in different circumstances. Moreover, the models that were used to research the impacts of certain circumstances can not incorporate the complexity of practical situations. For example, researchers often assume simple supply chains with a single retailer, a single manufacturer, a single supplier etcetera. Supply chains in practice, are often complex networks of many different companies.

In practice, not one situation is the same and theoretical models are generally much simplified. Therefore, the question raises what the value is of theoretical research on the bullwhip effect. Exact numbers or formula's are in my opinion only useful to compare different scenarios in certain situations. Chen et al. [6] derived a lower bound on the order rate variance ratio. In practical situations, this lower bound will often not be correct, because it heavily depends on the demand model, the inventory policy, forecasting technique and other assumptions. However, it gives an indication of the impact of a change, for example a change in the lead time or an other parameter. My opinion is that theoretical research on the bullwhip effect leads to more insight in the process and makes managers more aware of the problem.

Bibliography

[1] Kurt Salmon Associates. Efficient consumer response: Enhancing consumer value in the grocery industry. *Kurt Salmon Associates*, 1993.

[2] J. D. Blackburn. The quick-response movement in the apparel industry: a case study in time-compressing supply chains. *Time Based Competition: the Next Battleground in American Manufacturing*, 1991.

[3] G. P. Cachon and M. Fisher. Supply chain inventory management and the value of shared information. *Management Science*, 46(8), 2000.

[4] S. Cannella and E. Ciancimino. On the bullwhip avoidance phase: supply chain collaboration and order smoothing. *International Journal of Production Economics*, 48(22), 2010.

[5] D.C. Chatfield, J. G. Kim, T. P. Harrison, and J. C. Hayya. Impact of stochastic lead time, information quality, and information sharing: a simulation study. *Production and Operations Management*, 13(4), 2004.

[6] F. Chen, Z. Drezner, J. K. Ryan, and D. Simchi-Levi. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, 46(3), 2000.

[7] J. Dejonckheere, S. M. Disney, M. R. Lambrecht, and D. R. Towill. Measuring and avoiding the bullwhip effect: a control theoretic approach. *European Journal of Operational Research*, 147(3), 2003.

[8] S. M. Disney and A. Potter. Bullwhip and batching: An exploration. *International Journal of Production Economics*, 104(2), 2006.

[9] S. M. Disney and D. R. Towill. A discrete transfer function model to determine the dynamic stability of a vendor managed inventory supply chain. *International Journal of Production Research*, 40(1), 2002.

[10] S. M. Disney and D. R. Towill. On the bullwhip and inventory variance produced by an ordering policy. *Omega, the International Journal of Management Science*, 31(3), 2003.

[11] S. M. Disney et al. Controlling bullwhip and inventory variability with the golden smoothing rule. *European Journal of Industrial Engineering*, 1(3), 2007.

[12] J. W. Forrester. *Industrial dynamics.* MIT Press, Cambridge, MA, 1961.

[13] D. Heyman and M. Sobel. *Stochastic models in Operations Research Vol. 1.* McGraw-Hill, New York, 1984.

[14] M. Holweg, S. Disney, J. Holmstrom, and J. Smaros. Supply chain collaboration: making sense of the strategy continuum. *European Management Journal*, 23(2), 2005.

[15] T. Hosoda, M. M. Naim, S. M. Disney, and A. Potter. Is there a benefit to sharing market sales information? linking theory and practice. *Computers and Industrial Engineering*, 54(2), 2008.

[16] T. Kelepouris, P. Miliotis, and K. Pramatari. The impact of replenishment parameters and information sharing on the bullwhip effect: a computational study. *Computers and Operations Research*, 35(11), 2004.

[17] H. L. Lee, V. Padmanabhan, and S. Whang. The bullwhip effect in supply chains. *Sloan Management Review*, 38(3), 1997.

[18] H. L. Lee, V. Padmanabhan, and S. Whang. Information distortion in a supply chain: The bullwhip effect. *Management Science*, 43(4), 1997.

[19] R. Metters. Quantifying the bullwhip effect in supply chains. *Journal of Operations Management*, 15, 1997.

[20] J. K. Ryan. *Analysis of Inventory models with limited demand information.* PhD thesis, Department of Industrial Engineering and Management Science, Northwestern University, Evanstan 2, 1997.

[21] J.S. Shang, S. Li, and P. Tadikamalla. Operational design of a supply chain system using the taguchi method, response surface methodology, simulation, and optimization. *International Journal of Production Research*, 42(18), 2004.

[22] J. D. Sterman. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3), 1989.