Invited Review

The bullwhip effect: Progress, trends and directions

Xun Wang1, Stephen M. Disney*

Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff CF10 3EU, United Kingdom

Abstract

The bullwhip effect refers to the phenomenon where order variability increases as the orders move upstream in the supply chain. This paper provides a review of the bullwhip literature which adopts empirical, experimental and analytical methodologies. Early econometric evidence of bullwhip is highlighted. Findings from empirical and experimental research are compared with analytical and simulation results. Assumptions and approximations for modelling the bullwhip effect in terms of demand, forecast, delay, replenishment policy, and coordination strategy are considered. We identify recent research trends and future research directions concerned with supply chain structure, product type, price, competition, and sustainability.

© 2015 The Authors. Published by Elsevier B.V.

1. Introduction

The bullwhip effect is one of the most popular and celebrated concepts in the operations management/research field. The term ‘bullwhip’ was coined to describe the effect by which slow moving consumer demand creates large swings in production for the suppliers at the other end of the supply chain. This is analogous to the handle of the bullwhip causing a loud crack at the popper. The bullwhip effect is sometimes referred to as ‘demand amplification’, ‘variance amplification’ or the ‘Forrester effect’. This effect becomes significant when the cost from fluctuations in production/ordering outweighs the cost of holding inventory. Over the years, evidence has suggested that bullwhip costs play a pivotal role in some businesses. Bullwhip costs can be associated with setting up and shutting down machines, idling and overtime in the workload, hiring and firing of the workforce, excessive upstream inventory, difficulty in forecasting and scheduling, systems nervousness, and poor supplier/customer relationships, amongst other consequences.

The bullwhip effect also has a close link with the philosophy of lean production (Ohno, 1988). Mura—the waste of unevenness—is the failure to smooth demand and is recognised as the root cause of both Muda (the seven lean wastes) and Muri (the waste of overburden). Indeed Ohno (1988) discusses the benefits of bullwhip avoidance:

“The slower but consistent tortoise causes less waste and is much more desirable than the speedy hare that races ahead and then stops occasionally to doze. The Toyota Production System can be realized only when all the workers become tortoises.”

Since the 1990s, a large amount of literature on the bullwhip effect and its various proofs, interpretations, and remedies has emerged and continues to grow. A search in the Web of Science with the keyword ‘bullwhip effect’ returns 582 papers, highlighting a strong academic interest. This review summarises the achievements and findings of the past 20 years regarding the bullwhip effect and identifies possible future research directions. However, we do not confine our review solely to this 20-year interval, since much research on the macroeconomic, microeconomic and operational levels have a longer history and we include older contributions when they provide important contextual information.

Due to the various orientations and disciplines of research papers under this topic, a statistical systematic review is not appropriate. Rather, this review is narrative in nature as this allows more flexibility. We started collecting papers with a keyword search from the databases of Web of Science, EBSCO, and ScienceDirect. We have searched with the keywords ‘bullwhip effect’, ‘demand amplification’ and ‘variance amplification’ in the title, abstract and keyword sections. We also conducted a careful citation search both prospectively and retrospectively. This led to 455 papers being reviewed and over 150 articles being cited in this paper. Certainly, the page limitation has restricted us from citing all the papers we collected. However we have tried to include all the papers that are both relevant and significant. We have attempted to be as inclusive as possible but we are aware that we may have missed some important contributions on the topic. For the sake of readability and succinctness, we have avoided deep technical details. However, we do discuss issues such as assumptions and cost criteria since they are an essential basis for critical evaluation.
Several reviews have been devoted to the bullwhip effect. For example, Geary, Disney, and Towill (2006) classified five routes to increase our knowledge of bullwhip effect and 10 principles to reduce it. Miragliotta (2006) reviewed bullwhip research in three categories; empirical assessment, causes, and remedies, and then proposed a new taxonomy to model this problem. Giard and Sali (2013) categorised 53 bullwhip papers within 13 coordinates, including modelling approaches, demand models, measures, and causes. Other reviews are more conceptually oriented, attempting to offer a new perspective on bullwhip (Towill, Zhou, & Disney, 2007).

Some reviews are not solely confined to the bullwhip effect, but also cover other supply chain modelling issues (Beamon, 1998; Min & Zhou, 2002; Sarimveis, Patrinos, Tarantilis, & Kiranoudis, 2008). These papers assess general supply chain modelling methods, in which most bullwhip models have been categorised as stochastic analytical/simulation models. The bullwhip effect has also been frequently mentioned in review papers dedicated to other topics, such as information sharing (Sahin & Robinson, 2002) and reverse logistics (Govindan, Soleimani, & Kannan, 2014). As a result of this review, we identify research trends which require innovative models for bullwhip effect and we propose possible directions for future research.

The paper is organised as follows. In Section 2 we provide background information; a brief research history of the bullwhip effect, costs that are typically considered to be relevant to bullwhip, and bullwhip measures. In Section 3 we critically review research based on empirical and experimental methodologies. Bullwhip modelling specifics are dealt with in Section 4. Section 5 concludes and proposes future research directions by identifying a number of trends in bullwhip research that have emerged during recent years.

2. Background

2.1. A brief history of bullwhip research

The term bullwhip effect was first coined by Procter & Gamble (P&G) in the 1990s to refer to the order variance amplification phenomenon observed between P&G and its suppliers. Interestingly, a similar phenomenon between P&G and its wholesalers has been documented during 1910s (Schisgall, 1981). This effect is commonly observed in almost every industry. Fig. 1 shows four sample time series of empirical bullwhip evidence we have observed in real supply chains. We prefer not to identify the specific data sources, but we do highlight the industry in Fig. 1 where the increase of variation between demand and production orders (or shipments) can be clearly seen. Evidence can also be found in pasta (Hammond, 1994), automotive (Taylor, 1999), and retail (Lai, 2005) industry case studies.

Forrester (1961) first formalised the variance amplification effect using the ‘industrial dynamics’ approach. He later established a simulation experiment mimicking the decision making behaviour in supply chains—the famous ‘Beer Game’. Amplification turned out to be inevitable. Sterman (1989) published 20 years of data from the game attributing the amplification to the tendency that players overlook the inventory-on-order (the orders placed but not yet received), a cause of amplification known as ‘irrational behaviour’.

Please cite this article as: X. Wang, S.M. Disney, The bullwhip effect: Progress, trends and directions, European Journal of Operational Research (2015), http://dx.doi.org/10.1016/j.ejor.2015.07.022
The important work of Lee, Padmanabhan, and Whang (1997) not only brought the term bullwhip effect to widespread academic attention, but also proposed an additional four causes to the problem where players are assumed to behave completely rationally. These are demand signal processing, batch ordering, price fluctuation and shortage gaming.

2.2. Bullwhip related costs

Cost structures determine the optimal replenishment strategy. The balance between production/ordering related cost and inventory related cost governs whether one should amplify or smooth production. The production smoothing hypothesis (Holt, Modigliani, Muth, & Simon, 1960) assumes that production fluctuations increase the operational cost to the manufacturer by inducing excess machine setup, idle time and workforce hiring/firing. In order for production smoothing to be efficient, a quadratic—or more generally, a convex—cost on production/order quantity, or a cost on production/order changes must be present. This is a standard assumption in the bullwhip literature and is different from the assumption of a fixed ordering cost in other inventory management literature (see e.g. Clark & Scarf, 1960).

Inventory/backlog cost can either be assumed to be quadratic or piecewise linear to the inventory quantity. In many applications, inventory cost increases with the standard deviation of the inventory levels, as shown by the news-vendor model. There is also a trade-off between inventory cost and production cost, due to the stabilizing effect of inventory (Baganha & Cohen, 1998; Disney, Towlil, & van de Velde, 2004). Chen and Samroengraja (2004) showed that when the cost function is concave, the replenishment policy that minimises order fluctuations is not necessarily the one that minimises total cost.

2.3. Measures of bullwhip

The way we measure the bullwhip influences the quantification and evaluation of the effect. By definition bullwhip is the amplification of order volatility along the supply chain. This volatility can be measured by the coefficient of variation, variance, or standard deviation. As the variance frequently emerges naturally from mathematical investigations, it is often convenient to measure bullwhip by comparing the variance between demand and orders. Such comparison can be made by either a ratio or a difference, where amplification (smoothing) is indicated by a ratio larger (smaller) than one, or a difference greater (less) than zero (Cachon, Randall, & Schmidt, 2007). Due to data availability, some empiricists use alternatives such as production quantity, sales and shipments which are easier to observe than orders and demand (Blinder & Maccini, 1991). Under non-stationary demand it is necessary to perform difference operations on the time series. That is, to measure bullwhip by the variance of order changes instead of the variance of orders itself (West, 1986). Alternatively one may compare the difference between order variances and demand variances which has been proved to be finite (Gaalman & Disney, 2012). If the inventory system is to be modelled linearly, then the variance ratio is convenient because it coincides with an engineering concept called the noise bandwidth, a concept with an established theoretical basis (Aström, 1970).

With the above mentioned bullwhip measures, production smoothing and bullwhip are two opposite phenomena, indicating attenuation and amplification of order fluctuation. However, Bray and Mendelson (2015) argued that if we alter the measurement of bullwhip effect from an increase of order variability to that of order uncertainty, then these two effects may coexist.

3. Empirical and experimental research in bullwhip

We categorise the methodologies adopted in bullwhip research into: empirical, experimental, analytical and simulation-based. In empirical studies on bullwhip, historical data on demand, sales, shipment and production is collected and analysed. This sometimes comes with detailed background information of the company or supply chain under investigation. This is not only efficient in detecting bullwhip, but also enables one to pinpoint underlying causes; test inductive hypotheses and corollaries; and to measure the performance of implemented remedies. Nonetheless, observations and conclusions are often mixed and sometimes contradictory.

Experimental research uses laboratory experiments and management games to examine factors and mechanisms that affect bullwhip. Usually they focus on the behavioural, psychological and cognitive aspects of decision-makers with regard to their forecasting, replenishment, or capacity-setting behaviours. The experimental method allows theories to be tested in an isolated and controlled environment, reducing the impact of exogenous disturbances. When properly designed the experiments are also suitable for pedagogical purposes.

Mathematically, this is achieved by precisely quantifying the bullwhip effect and its causes, to predict the response of the system to various types of disturbances, and offer guidelines for prevention and elimination. It is often beneficial to simplify the model and seek rigorous analytical insights. On the other hand, simulation offers a chance to tackle more realistic bullwhip problems numerically and computationally when the complexity of the mathematical model is beyond our analytical capability. The major challenge is to reasonably simplify and abstract the real problem into a mathematical one.

3.1. Empirical studies

Some of the first examples of demand amplification were found by economists. They termed this the ‘investment accelerator effect’, referring to the phenomenon that a given change in demand for consumer goods results in a more severe change in demand for capital goods. This effect is considered to be an important cause of business and economic cycles (Clark, 1917; Mitchell, 1913; Samuelson, 1939). Vaile, Grether, and Cox (1952) reported that the production of consumer goods is estimated to have fallen from an index of 100–80 during the 1929–1932 depression, while that of all capital equipment fell from 100 to 35. Such an effect is sufficient to generate demand amplification. Bishop (1984) provided evidence of the amplification effect between fossil-fuel demand and turbo machinery demand. The cause was attributed to the time delay in capacity investment and the speed of machine wear-out (Hicks, 1950).

In 1960 Holt et al. (1960) proposed the production smoothing model assuming that rational decisions regarding production quantities would lower costs by levelling production, with inventory being used as a buffer. Efforts have been made to optimise this model under various assumptions (Gaalman, 1978; Schneeeweiss, 1974; Zangwill, 1966). This idea theorises that one should be able to observe that: (1) production is smoother than consumption; and (2) inventory is negatively correlated to consumption. Quite contrarily, many empirical studies have found amplification between retail sales and production orders, as well as positive correlation between demand and inventory (Blanchard, 1983; Blinder, 1986; Blinder & Maccini, 1991; West, 1986). These can be viewed as early examples of the bullwhip effect in the production echelon, an effect that was then termed ‘excess volatility’.

Several explanations for this excess volatility have been proposed. First, it has emerged that data aggregation plays a vital role in the measurement of amplification. Data used in production smoothing research are often extracted from open-access statistical reports of industry scale on a quarterly or annual basis. Second, in these reports, entries for inventories and sales (shipments) are usually recorded in monetary terms, and rarely on an SKU basis. Therefore, price and seasonality are often commingled with quantity data and deflation and deseasonalisation techniques are required (Allen, 1999). Others
advocate that production smoothing is more easily found when demand has a predictable seasonal component (Ghali, 1987). Wang (2002) extracted data from 46 product items and pinpoints price variation as a contributing factor to the production smoothing phenomenon. However, incorporation of price and seasonal fluctuation does not always generate results in support of production smoothing (Miron & Zeldes, 1988). Other factors, such as non-convex costs, random prices, variable target stocks and lost sales have also appeared in explanations of the excess volatility (Blinder, 1986; Kahn, 1992; Milne, 1994; Ramey, 1991; West, 1986).

In the majority of surveys, demand amplification appears to be dominant. Ghali (2003) showed that production smoothing can be found only in a small number of industries where seasonality is stable and inventory holding cost is low. In 75 industries, Cachon et al. (2007) observed that 61 exhibited bullwhip when seasonality was removed, but only 39 when not. Similar findings have been reported by Bray and Mendelson (2012), on the basis of firm-level, rather than industry-level, data. In 31 firms under investigation, 30 and 26 exhibited the bullwhip effect with and without seasonality removed respectively. Other studies, including those conducted by Fair (1989), Ghali (1974) and Krane and Braun (1991), were in favour of the smoothing hypothesis.

Operations management empiricists have also examined how the amplification grows along the chain. Baganha and Cohen (1998) observed that bullwhip effect appears in the wholesaler’s echelon, and argued that the wholesaler’s inventory acts as a stabiliser in the chain. Using U.S. industry-level data, Cachon et al. (2007) also found that bullwhip primarily appears in the wholesaler, rather than in the retailer or manufacturer, echelon. Dooley, Yan, Mohan, and Gopalakrishnan (2010) studied the bullwhip effect during the 2007–2009 recession and concluded that retailers responded to market changes rapidly and adaptively, whereas wholesalers responded late and drastically.

It is also believed that smoothing/amplification behaviour may vary among different nations and cultures. Mollick (2004) described evidence of production smoothing in the Japanese automotive industry, where the production smoothing is more common due to the prevalence of Heijunka (levelling) and Just-In-Time manufacturing strategies. Shan, Yang, Yang, and Zhang (2014) studied the bullwhip effect in China, finding that bullwhip was gradually being reduced.

### 3.2. Experimental research

The experimental approach was pioneered by the seminal paper of Sterman (1989), who documented a role-playing game for inventory management called the ‘Beer Game’. This later became the standard experimental framework to study supply chain dynamics. Participants in the game act as firms along a beer distribution chain and make ordering and production decisions. Amplification as large as 700% was commonly observed in the four echelon setting. Later variations of the game introduced: random and seasonal demand; sharing of demand and inventory information (both on-hand and on-order); training before the actual experiment; and trust between participants. The bullwhip effect persists in most cases (Croson & Donohue, 2005, 2006; Croson, Donohue, Katok, & Sterman, 2014; Nienhaus, Ziegenbein, & Schoensleben, 2006; Wu & Katok, 2006).

Sterman (1989) understood the order volatility from the perspective of bounded rationality and sub-optimal decisions. By analysing Beer Game results he discovered that most participants tend to overlook the on-order inventory (the supply-line or work-in-process) when making replenishment decisions. This phenomenon repeatedly occurred in subsequent experimental studies. This underweighting does not improve when: the supply line is made visible (Wu & Catok, 2006); demand is known and stationary (Croson & Donohue, 2006); or even when demand is known and constant (Croson et al., 2014). Moreover, there is evidence to suggest that some participants choose a positive feedback strategy, i.e., they restrain their orders when large stock-outs have built up (Delhoum & Scholz-Reiter, 2009).

Another explanation for the amplification phenomenon is the objective of the game. Players are tasked with minimising inventory holding and backorder costs. A rational player should choose a base stock policy to minimise costs and simply ‘pass on orders’. This clearly does not smooth orders. Cantor and Katok (2012) introduced a cost for production and order changes, and found that production is smoothed when demand is seasonal, and that the smoothing behaviour is more eminent when the production change cost is high.

### 3.3. The existence of bullwhip effect

In this subsection we revisit the debate over the existence of bullwhip effect and production smoothing, which are both extensively verified empirically and experimentally. The question that naturally arises is, if variable production is indeed costly, why does a homo economicus choose to bullwhip? This question was posed by Cachon et al. (2007), which Sterman (1989) explained as an irrational mistake and Lee et al. (1997) as a rational choice.

Compared with optimizing within a pre-determined cost function, it is much harder to observe the impact of bullwhip on the profitability of a company. Many of the consequences of the bullwhip effect are hard to quantify economically. These include the cost of hiring, firing, learning and training, overtime and idling, as well as the impact of the increased demand variability on the upstream suppliers. In this regard, both Sterman’s (1989) and Lee et al. (1997) explanations are inadequate since the cost assumptions in both approaches inherently induce amplification. Hence a lot of questions remain open regarding the emergence of bullwhip in real supply chains. Are production costs convex? Do decision makers recognise this? What would cause them to behave differently? How should they behave?

Consequently, we are still uncertain of the precise circumstances where we can detect bullwhip. For instance, empirical studies have suggested that it appears mostly in the wholesaler’s echelon, which contradicts the ‘continual amplification’ predicted by most of the cascading theoretical models. Also, production smoothing is often observed when demand is highly seasonal. This suggests that, despite its undeniable existence, the bullwhip effect may not be universal and may be explained by the different cost incentives that firms face. Research in this direction can benefit from empirical, experimental and analytical approaches to identify the real cost structure as well as the typical and optimal dynamic response. Specifically, case studies and empirical analysis based on firm-level rather than industry-level data have the ability to offer more insights on the incentives of smoothing and amplification.

### 4. Elements in bullwhip modelling

The conventional technique to examine the bullwhip effect analytically is to model supply chain participants as a dynamical inventory system. The impact of elements such as demand, delay, forecasting policy, ordering policy and information sharing mechanisms can be investigated. Some of these factors are deemed exogenous to the decision maker, some endogenous. All of them could have either positive or negative impacts on demand amplification. Fig. 2 provides a sketch of the roles of these elements.

#### 4.1. Demand

Chen, Dresner, Ryan, and Simchi-Levi (2000a) identified that bullwhip is, at least partly, due to the unpredictability of demand, lead-times and the need to forecast future demand. Since then these causes have received a large amount of research attention.
Most research in this direction has assumed that demand is a stochastic process. The simplest demand model is an independently and identically distributed (i.i.d.) Gaussian white noise process (Deziel & Eilon, 1967). This model has some mathematical advantages, but may be an over-simplification as it overlooks temporal correlation in the demand signal. Demand correlation can be accounted for with auto-regressive integrated moving average (ARIMA) models (Box & Jenkins, 1970). The first order auto-regressive demand model, AR(1), has perhaps been the most frequently adopted (Chen et al., 2000a; Lee et al., 1997, 2000, amongst others). This demand model has only one parameter, so it is easy to observe the impact of autocorrelation without sacrificing too much tractability. It has been shown that positive (negative) correlation in demand increases (mitigates) bullwhip (Duc, Luong, & Kim, 2008a) and that there is an auto-regressive parameter with maximal bullwhip for a given lead time (Babai, Boylan, Syntetos, & Ali, 2015). For positive correlation, bullwhip increases with the lead time until it reaches an upper bound (Luong, 2007).

More complex ARIMA models for demand have also been studied: AR(2), AR(p) (Luong & Pien, 2007); ARMA(1,1) (Alwan, Liu, & Yao, 2003); ARMA(2,2) (Gaalmann & Disney, 2009); and ARMA(p,q) (Gaalmann, 2006). For non-stationary ARIMA models, since the variance of non-stationary time series is no longer finite, alternative bullwhip measures and approaches have to be adopted e.g. the variance of non-stationary ARIMA processes. The bullwhip effect problem under MMFE demand was considered in Chen and Lee (2009).

4.2. Forecasting

A wide range of forecasting methods have been investigated in the bullwhip literature. Chen et al. (2000a) and Duc et al. (2008a) studied the moving average (MA) forecasting method, while Chen, Ryan, and Simchi-Levi (2000b) and Dejonckheere, Disney, Lambrecht, and Towill (2003) investigated the simple exponential smoothing (SES) method. These are both user-friendly forecasting techniques that have been widely adopted in industry. Minimum mean squared error (MMSE) forecasting minimises the expectation of the squared forecast error. It is capable of achieving the highest possible accuracy for the specified demand process, and often used as a benchmark (Alwan et al., 2003; Hosoda & Disney, 2006a; Zhang, 2004a).

However, the utilisation of MMSE forecasting is based on the often unrealistic assumption that the demand structure can be specified, is known to the forecaster, and is constant over time. Therefore, some authors (e.g. Aviv, 2003) have adopted the Kalman filter approach. This is essentially a recursive algorithm that converges to the MMSE forecast over time. The impact of more sophisticated forecasting methods such as Holt’s, Brown’s and Damped Trend Forecasting was discussed by Wright and Yuan (2008) and Li, Disney, and Gaalman (2014). These forecasting techniques are designed for seasonal and trended demand.

Another interesting topic is the relationship between forecast accuracy and total cost. Zhang (2004a) suggested that MMSE forecasting minimises inventory-related cost. This was supported by Hussain et al. (2012) in a simulation study. However, according to some empirical (Plores, Olsen, & Pearce, 1993) and analytical research (Hosoda & Disney, 2009), the most accurate forecasting does not always result in an optimal supply chain when local bullwhip or global inventory costs are taken into account (Disney, Lambrecht, Towill, & Van de Velde, 2008; Gaalmann, 2006; Gaalmann & Disney, 2006; Gaalmann & Disney, 2009).

In Fig. 3, we summarise the existing research on demand evolution, i.e., the structure of the replenishment order series generated by different ARIMA demand models and forecasting methods, sometimes referred to as ‘demand propagation’. This knowledge is useful for quantifying the bullwhip effect because once an ARIMA process is specified, it can readily be transformed into an infinite MA process, from which variances are easily calculated.
by Sterman (1989) to depict the misperception of delays in the Beer Game. It is also known as ‘proportional feedback control’. Consequently, the effort to design a satisfactory feedback parameter has shed light on how lead-time misperception affects the bullwhip effect. The proportional feedback control technique was introduced by Magee (1956) and further developed by Deziel and Eilon (1967) and Towill (1982). The automatic pipeline, inventory and order-based production control system (APIOPCS) proposed by John, Naim, and Towill (1994) is mathematically equivalent to Sterman’s (1989) ‘anchoring and adjustment heuristic’.

When the system is linear and time invariant and the cost function is quadratic, the optimal policy is known to be linear and can be solved with the Riccati equation (Sethi & Thompson, 2000). Deziel and Eilon (1967) proposed the first linear proportional production control policy where the same feedback parameter is assigned to both the inventory and pipeline levels. Proportional control has the capability to reduce order variance (Chen & Disney, 2007; Lin, Wong, Jang, Shieh, & Chu, 2004), at the expense of increased inventory variance and reduced customer service (Disney et al., 2008; Hosoda & Disney, 2006b; Jakšić & Rusjan, 2008). An optimal parameter setting is given by Disney et al. (2004) for the Deziel–Eilon model. Interestingly, when order and inventory variances are equally weighted in the cost function and demand is i.i.d., the golden ratio describes the optimal controller. General guidance on tuning the feedback parameters is given by Balakrishnan, Geunes, and Pangburn (2004), Papangnou and Halikias (2008), Graves, Kletter, and Hetzel (1998) and Boute and Van Miegham (2015) describe other proportional ordering policies.

**Batched policies**: Ordering in batches allows for economies of scale in ordering, set-up or transportation. Under these policies, there is a minimum ordering quantity of Q products, which leads to an impulsive order process. It is generally believed that a smaller batch size helps to stabilise orders and to reduce operational cost (Burbridge, 1961; Caplin, 1985; Holland & Sodhi, 2004; Lee et al., 1997; Wangphanich, Kara, & Kayis, 2010). However, if the batch size is a multiple of average demand, then reducing the batch size may not be necessary (Li & Brdihan, 2008; Potter & Disney, 2006).

**Aggregation issues**: The problem of product/location aggregation arises when a supplier maintains multiple retailers, or distribution centres in different locations, or by manufacturing different products on the same line. This problem has been investigated under (s,S) (Caplin, 1985; Kelle & Milne, 1999), (Q,T) (Cachon, 1999; Lee et al., 1997) and base stock (Sucky, 2009) policies. The problem of temporal aggregation arises when time series data has to be summed on a periodic, non-overlapping basis for review and decision-making purposes. Quarterly financial reports and weekly replenishment orders are examples of temporal aggregation. It is found that aggregation in both location and time has a masking effect on the bullwhip effect, i.e., bullwhip decreases with the aggregation period but cannot be fully eliminated (Chen & Lee, 2012; Noblebsse, Baute, Lambrecht, & van Houdt, 2014).

### 4.5. Information sharing

It has been advocated that the bullwhip effect can be mitigated by information sharing. Lee and Whang (2000) summarised the common schemes for sharing information on inventory levels, sales data, sales forecast, order status and production/delivery schedules. This information can be shared both upstream and downstream. More advanced integration allows supply chain members to collaborate with, or consign planning and replenishment decisions to, their peers. We review research on two supply chain cooperation mechanisms; demand information-sharing, and vendor-managed inventory (VMI).

**Demand information sharing**: Information sharing is a term used to describe the case where end consumer demand is communicated to all members of the supply chain, who then use that information in their forecasts, despite being required to deliver what their immediate customer requests. Lee et al. (1997) proposed demand information sharing as a countermeasure to the bullwhip effect generated by demand signal processing. It has since become the most commonly investigated coordination mechanism. Theoretically, the effectiveness of information sharing in reducing bullwhip has also been demonstrated with: the base stock policy (Chatfield et al., 2004; Dejonckheere, Disney, Lambrecht, & Towill, 2004); batch ordering (Hussain & Drake, 2011); correlated demand (Gaur, Giloni, & Seshadri, 2005; Lee, So, & Tang, 2000); price fluctuations (Gavirneni, 2006; Ma, Wang, Che, Huang, & Xu, 2013); and in reverse supply chains (Adenso-Díaz, Moreno, Gutiérrez, & Lozano, 2012). Under certain circumstances, order variance increases linearly with information sharing and exponentially without (Dejonckheere et al., 2004; Kim et al., 2006). Several factors influence the potential benefit of information sharing, including demand patterns (Steckel et al., 2004) and lead-times. For instance, information sharing is more beneficial when demand is highly correlated or highly variable, or when the lead-time is long (Babai et al., 2015; Lee et al., 2000). The direction of information sharing (i.e., upstream or downstream) also affects its benefit (Yao & Zhu, 2012).

From experimental and analytical evidence, some authors have found that information sharing alone cannot eliminate the bullwhip effect (Chen et al., 2000a; Croson & Donohue, 2006; Ouyang, 2007; Sodhi & Tang, 2011). In addition, since information systems often require substantial investment, comparisons have been made regarding the performance of information sharing and alternative bullwhip reduction schemes, such as advanced demand information (Ouyang & Daganzo, 2006; Raghunathan, 2001) and lead-time reduction (Agrawal et al., 2009; Chen and Lee, 2009).

**Vendor managed inventory**: Implementation of VMI requires sharing of both demand and inventory information. Under a typical VMI agreement, the supplier puts the customer’s inventory under its surveillance and automatically replenishes it. It is believed that VMI removes decision echelons in the supply chain and reduces the risk of information distortion and the amplification effect (Cannella & Ciancimino, 2010; Disney & Towill, 2003; Xu, Dong, & Evers, 2001). Dong, Dresner, and Yao (2014) presents an empirical study of the benefit of VMI based on item-level data finding that VMI benefits downstream firms by reducing inventory and stock-outs, while upstream firms benefit from bullwhip reduction.

## 5. Opportunities for future research—what can we expect in the next decade?

Interest in the bullwhip phenomenon began almost a century ago in the macroeconomics literature, and has thrived at the microeconomic and operational level in the last 20 years. Extensive studies over the last two decades have profoundly changed our understanding of this phenomenon, and we have reached some consensus, or at least dominating opinions, which we now summarise:

- Bullwhip can be observed at the industry-, firm-, and product-level, in various types of supply chains.
- Bullwhip can be induced by both rational and irrational decision-making behaviour.
- Under certain circumstances, bullwhip can be reduced or even eliminated.
- More accurate forecasts, smaller batch sizes and shorter lead-time help to reduce bullwhip.
- Supply chain integration, collaboration, information transparency and centralised decisions are also beneficial.

We have seen how these opinions are formed and the traditional approaches are used to explain and solving the bullwhip problem. We have also discovered that recently some innovative models and techniques have been developed to relax assumptions and to reveal richer characteristics of supply chains, which we summarise as (see Fig. 4):
5.2. Bullwhip in service chains due to conceptual and technical challenges. These attempts are not new but they remain largely underdeveloped.

5.3. Bullwhip with price consideration

The negative correlation between price and demand naturally means that price variability results in demand variability. Lee et al. (1997) adopted a simple model to illustrate this effect. Özelkan and Çakanyıldırım (2009) studied financial flows in a game theoretical two-echelon supply chain model. They found that retail prices were more variable than wholesale prices, a phenomenon known as ‘the reverse bullwhip of prices’. Others have attempted to investigate this problem more from an operations perspective, ignoring the economic consequences. Among these are Zhang and Burke (2011), who showed that introducing price fluctuations can either exacerbate or mitigate the bullwhip effect, based on the auto- and mutual-correlation between price and demand. Recently Sodhi, Sodhi, and Tang (2014) incorporated a discretely distributed stochastic price into the economic order quantity model. They showed that the bullwhip effect persists and is positively related to the variance of price. The effect of price fluctuation is particularly significant when the fixed ordering cost is small.

Lee et al. (1997) suggested that price stabilization or everyday low price (EDLP) helps to mitigate this problem. This strategy has been implemented in several retail chains, such as ASDA and Walmart. However, the validity of this measure remains questionable. Some have suggested that EDLP cannot decrease order variability and firms would not choose to reduce bullwhip-induced costs when it jeopardizes more important objectives such as market share, total revenue, and service level (Su & Geunes, 2012). Alternatively Gavirneni (2006) and Hamister and Suresh (2008) argued that a fluctuating pricing policy may lower demand volatility and improve profitability. Research on the influence of prices on bullwhip requires models that incorporate price setting and negotiation processes, dramatically increasing the complexity of the model.

5.4. Bullwhip with resource competition

Limited supply induces order variation, since customers order more than what they require to ensure that their needs are met; a phenomenon otherwise known as ‘order inflation’ or ‘rationing and gaming’. Lee et al. (1997) attempted to explain this effect using a
one-supplier-multiple-retailer newsvendor model. Supplies were allocated proportionally, and it was shown that when supply is insufficient, the optimal ordering quantity is always larger than the normal newsvendor quantity. This study was followed by several papers discussing which allocation mechanisms or policies will prevent retailers from inflating their orders (a.k.a. ‘strategy-proof’ or ‘truth-inducing’ allocation mechanisms). These allocation rules can be categorised into two groups:

- Individually responsive (IR)—if fulfilment strictly increases with orders;
- Individually unresponsive (IU)—the share of capacity is predetermined by the supplier. If a retailer orders less than the share, the complete order is received, and the remaining capacity is allocated among the other retailers in some manner.

Table 1 summarises commonly investigated allocation policies.

<table>
<thead>
<tr>
<th>Policy</th>
<th>Category</th>
<th>Rule</th>
<th>Research works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportional</td>
<td>IR</td>
<td>Stock is allocated according to the proportion of customers’ orders</td>
<td>Lee et al. (1997)</td>
</tr>
<tr>
<td>Linear</td>
<td>IR</td>
<td>All customers share the same amount of deficit</td>
<td>Cachon and Lariviere (1999a)</td>
</tr>
<tr>
<td>Uniform</td>
<td>IU</td>
<td>All customers have the same share, but those who order less than this share get what they ordered</td>
<td>Cachon and Lariviere (1995a)</td>
</tr>
<tr>
<td>Pareto</td>
<td>IU</td>
<td>Capacity is allocated to maximise supply chain profit assuming all retailers are truthful</td>
<td>Cachon and Lariviere (1999b)</td>
</tr>
<tr>
<td>Turn-and-earn</td>
<td>IU</td>
<td>Allocation according to past sales</td>
<td>Cachon and Lariviere (1999c)</td>
</tr>
<tr>
<td>Lexicographic</td>
<td>IU</td>
<td>Allocation according to some predetermined allocation priority sequence</td>
<td>Chen et al. (2013)</td>
</tr>
<tr>
<td>Competitive</td>
<td>IU</td>
<td>Allocation according to the proportion of customers’ optimal ordering quantity with unlimited capacity</td>
<td>Cho and Tang (2014)</td>
</tr>
</tbody>
</table>

Recycling and remanufacturing are the most commonly tackled only environmental issue in the bullwhip literature. Other green issues, such as pollution and carbon emission have received less attention. The social impact of bullwhip effect is notably missing from the literature. There are also theoretical assertions that bullwhip negatively affects the efficiency of workforce investment (Section 4.1), but the link between bullwhip and the welfare of employees and communities has not been studied previously.

5.6. Bullwhip as an extended concept

We see that the concept of bullwhip has been greatly extended since its introduction, from the amplification of material flow to a much larger set of amplification phenomena in cascading structures, including workload and price changes. In terms of cash flow, Tangscheeva and Prabhu (2013) and Chen, Liao, and Kuo (2013) proposed a ‘financial bullwhip effect’, measured by internal liquidity risk (in terms of bond yield spread) and the cash conversion cycle. The bullwhip term extends to refer to any kind of trend that is both repetitive and persistent. For instance, the term ‘green bullwhip effect’ is used by Lee, Klassen, Furlan, and Vinelli (2014) for the phenomenon that customer’s environmental requirements (in the form of specifications) become tighter and more stringent as they pass upstream. This leads to some exciting questions: What other forms can bullwhip take? So far we have seen bullwhip or bullwhip-like patterns in the forms of material flow, information flow, cash flow, work flow and even regulations. Is bullwhip universal in supply chain systems or a mere coincidence? Can it be observed in other cascading systems? Looking at amplification in different forms and other contexts will not only deepen our bullwhip understanding but will also reveal the true intricacy and beauty of the dynamics of supply chains.

Acknowledgements

We would like to thank the two anonymous referees for their constructive comments and our colleagues in Cardiff Business School for providing useful suggestions.

References


Please cite this article as: X. Wang, S.M. Disney, The bullwhip effect: Progress, trends and directions, European Journal of Operational Research (2015), http://dx.doi.org/10.1016/j.ejor.2015.07.022