Int. J. Production Economics 171 (2016) 311–320

Contents lists available at ScienceDirect

Int. J. Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Quantifying the bullwhip effect using two-echelon data: A cross-industry empirical investigation

Olov H.D. Isaksson [a,*], Ralf W. Seifert [a,b]

[a] École Polytechnique Fédérale de Lausanne (EPFL), College of Management of Technology, Chair of Technology & Operations Management, Station 5, 1015 Lausanne, Switzerland

[b] IMD, Chemin de Bellerive 23, P.O. Box 915, 1001 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 14 March 2013
Accepted 25 August 2015
Available online 3 September 2015

Keywords:
Bullwhip effect
Demand amplifications
Two-echelon data
Propensity score matching

ABSTRACT

The bullwhip effect denotes the phenomenon whereby demand variability is amplified from a downstream site (buyer) to an upstream site (supplier) in the supply chain. This paper contributes to the literature that empirically investigates the bullwhip effect by providing new evidence regarding its prevalence and magnitude. In contrast to previous work, we use a two-echelon approach, which allows us to observe variations at both the upstream and the downstream sites. By drawing on a financial accounting standard regarding information disclosure about major customers, we are able to link 5494 buyers and suppliers in the U.S. between 1976 and 2009. We merge this information with quarterly financial accounting data to form a sample of 14,933 buyer–supplier dyad observations. We correct for sample selection bias using propensity score matching and estimate the average bullwhip effect in our sample to be 1.90 (i.e. 90% demand variability amplification between echelons). A significant bullwhip effect is observed across industries (mining, manufacturing, wholesale and retail) and is supported by several robustness checks. We investigate and discuss how these results can be generalized beyond our sample.

1. Introduction

The bullwhip effect is said to occur when demand variability is amplified upstream in the supply chain. The first reference to this phenomenon is commonly attributed to Forrester's (1958) work on industrial dynamics, but evidence of its existence can be tracked as far back as the beginning of the 20th century (Geary et al., 2006). The bullwhip effect gained increased academic attention following the seminal paper by Lee et al. (1997b), and is now widely accepted as a source of great inefficiencies and costs in the supply chain (Metters, 1997). Firms that experience the bullwhip effect require higher capacity levels to cope with upswings in demand. In addition, stock-out costs are likely to increase during demand peaks, whereas warehousing and obsolescence costs increase on the downturn.

A number of case studies show evidence of the bullwhip effect in different industries. These case studies are very useful for studying the causes of the bullwhip effect and for identifying potential mitigating actions; however, the studies only provide limited evidence of the actual size and relative frequency of the effect. After all, any case study on the bullwhip effect is bound to involve a supply chain that bullwhips—i.e. there is a publication bias. Considering the central role of the bullwhip effect in supply chain management, there are surprisingly few empirical studies that try to quantify the effect on an aggregate level. Without rigorous empirical research, academics and managers must rely on anecdotal evidence and case studies. Large-scale empirical bullwhip estimates would therefore give a better understanding of the true importance of the phenomenon and help us to focus on research and attenuating actions where they are most needed. From an academic perspective, improved estimates would also have direct applications as parameters in mathematical inventory models. In this paper, our research objective is therefore to

Quantify the prevalence and magnitude of the bullwhip effect across industries, using multi-echelon data.

To the best of our knowledge, this is the first study to quantify the bullwhip effect using a multi-echelon approach. In the empirical operations literature, there are two central papers that set out to measure the bullwhip effect. Using industry level data, Cachon et al. (2007) detect the effect in the wholesale industry, but not in the manufacturing and retail industries. Bray and Mendelson (2012) use quarterly firm-level data, and find a mean bullwhip effect of 15.8% (demand variability amplification between echelons). Our study substantially differs from those of

* Corresponding author. Tel.: +41 21 693 24 72; fax: +41 21 693 00 20.
E-mail address: olov.isaksson@epfl.ch (O.H.D. Isaksson).

http://dx.doi.org/10.1016/j.ijpe.2015.08.027
0925-5273/

Cachon et al. (2007) and Bray and Mendelson (2012) in that it uses data that link buyers and suppliers (rather than relying on a single-echelon setting). These two previous studies use production at the focal firm as a proxy for outgoing orders to suppliers. Although this proxy yields great insights, our approach offers the advantage of observing variations at both the upstream and the downstream sites. We argue that the perspective of multiple echelons allows us to estimate the bullwhip effect in a way that better reflects its original definition. Furthermore, by building on the causes of the bullwhip effect, we provide arguments as to why a single-echelon setting is bound to yield downward-biased bullwhip estimates.

In our unique sample of 14,933 buyer–supplier dyads, we find strong support for the bullwhip effect across industries. We correct for sample selection bias using propensity score matching and estimate the mean amplification of demand variability between echelons to be 90% ($p < 0.001$). As noted above, prior studies find mixed evidence (Cachon et al., 2007) or much smaller bullwhips (Bray and Mendelson, 2012).

The remainder of this paper is structured as follows. In Section 2, we summarize the relevant literature on the bullwhip effect. Section 3 reviews the causes of the bullwhip effect and discusses them in the context of our multi-echelon approach. Our data set, linking buyers and suppliers, is introduced in Section 4. In Section 5 we develop a matching estimator to correct for sample selection bias and quantify the bullwhip effect across industries. In Section 6, we implement alternative specifications to check the robustness of our results. Section 7 analyzes our findings and discusses them in relation to previous work. We close with a discussion of research limitations and an outline of potential future research.

2. Literature review

2.1. The bullwhip effect

Most operations management students and scholars have been exposed to the bullwhip effect through The Beer Game and Hammond's (1994) case study of the Italian pasta producer Barilla. Since Hammond (1994), many other case studies have investigated and confirmed the bullwhip effect. Lee et al. (1997a) find evidence at a soup manufacturer and partly attribute it to price fluctuations. Fransoo and Wouters (2000) discuss how to measure the bullwhip effect and observe it in the supply chain of ready-made meals and salads. Anderson et al. (2000) find amplified upstream volatility in the machine tool industry. Lai (1992) shows order batching to be one of the leading causes of the bullwhip effect at a Spanish retailer. Terwiesch et al. (2005) study demand forecast sharing between buyers and suppliers, and observe amplification between demand for electronic devices and orders to semiconductor manufacturers. Wong et al. (2007) find unpredictable and seasonal demand to be a cause of the bullwhip effect in a toy supply chain. More recently, Zotteri (2013) analyzes fast moving consumer goods and cites price fluctuations and forward buys as the main drivers of demand variability amplification. Although a large overall bullwhip effect is observed, the effect varies substantially across product families.

The bullwhip effect has been well documented in the theoretical operations management literature. Following Lee et al. (1997b), various studies have greatly contributed to our understanding of the effects systemic causes, and in terms of identifying potential mitigating actions. Cachon (1999) finds that scheduled ordering policies can reduce supplier demand variability and lower total supply chain costs. Using a single-item inventory model, Graves (1999) observes how demand becomes increasingly variable upstream. Nevertheless, this amplification cannot be mitigated using information sharing. The opposite is observed by Chen et al. (2000), who model a two-stage supply chain and show that the bullwhip effect can be strongly reduced by centralizing demand information. These contradictory insights are ascribed by Gilbert (2005) to different modeling assumptions regarding the cause of the bullwhip effect. An array of studies investigating factors that can help reduce the bullwhip effect have followed. To name a few, there have been studies on vendor managed inventory (Disney and Towill, 2003), collaborative forecasting (Aviv, 2007), improved forecasting (Wright and Yuan, 2008), lead time reduction (Bayraktar et al., 2008), information sharing (Chen and Lee, 2009) and real-time visibility (RFID) (Bottani et al., 2010).

Similar research questions have also been investigated using experimental research methods. Croson and Donohue (2003, 2006) study the bullwhip effect from a behavioral perspective and find that it can be reduced with information sharing. The results of Steckel et al. (2004), by contrast, indicate that shorter cycle times can be beneficial, but that shared POS information does not necessarily help. Also using controlled experiments, Wu and Katok (2006) show that improved supply chain coordination can help alleviate the bullwhip effect.

The economics literature—rather than directly studying the bullwhip effect—has investigated whether firms smooth out their production using inventory as a buffer. This process is referred to as the production smoothing model and principally mirrors the trade-off between inventory holding costs and the cost of varying production. In light of this theory, economists have set out to empirically examine whether sales are more variable than production. Unexpectedly, they find little evidence of this phenomenon (Blanchard, 1983; Blinder, 1986; Kahn, 1987, 1992; Miron and Zeldes, 1988; Blinder and Maccini, 1991; Bivin, 1996). Gorman and Brannon (2000) argue that previous results are driven by the absence of seasonal adjustments, and they present evidence of production smoothing using seasonally adjusted data.

The lack of detailed data linking firms to supply chain networks is an obstacle to replicating case study findings on a larger scale. Nevertheless, in the empirical operations literature, there are two central papers that explore the bullwhip effect using a single-echelon approach (where production is used as a proxy for orders to suppliers). Cachon et al. (2007)—in contrast to Gorman and Brannon (2000)—argue and show that seasonal adjustments bias results towards a positive bullwhip effect. The authors use monthly U.S. Census industry-level data and obtain mixed results. With unadjusted data, the effect is found in the wholesale industry (net amplification of 14%) but not in the manufacturing and retail industries. Bray and Mendelson (2012) use a martingale forecast evolution model to decompose the bullwhip by information lead time. Applying their estimator to quarterly Compustat firm-level data, they find a mean bullwhip effect of 15.8% (demand variability amplification). As mentioned in Section 1, this study differs significantly from the studies of Cachon et al. (2007) and Bray and Mendelson (2012) in that it uses a multi-echelon—rather than a single-echelon—approach. In Section 3 we provide arguments regarding why the use of single-echelon data is bound to underestimate the bullwhip effect.

To conclude, there is a large body of research that provides evidence of the bullwhip effect and gives substantial insights into its causes and potential mitigating actions. Still, none of the studies mentioned above can be considered a measure of the overall prevalence and actual magnitude of the effect, given its definition as a multi-echelon phenomenon. This is the gap this research aims to fill.