Int. J. Production Economics 146 (2013) 543–549

Contents lists available at ScienceDirect

Int. J. Production Economics

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

Small sample uncertainty aspects in relation to bullwhip effect measurement

Erland Hejn Nielsen*

Department of Business Studies, Aarhus School of Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark

A R T I C L E I N F O

Article history:
Received 19 April 2010
Accepted 2 August 2012
Available online 13 August 2012

Keywords:
Bullwhip effect
Bullwhip measure
Bullwhip drivers
Small sample properties

A B S T R A C T

The bullwhip effect as a concept has been known for almost half a century starting with the Forrester effect. The bullwhip effect is “allegedly” observed in many supply chains, and it is generally accepted as a potential malice. Despite this the bullwhip effect still seems to be first and foremost a conceptual phenomenon. Some even simply denies that the phenomenon exists at all in practice. This of course makes it important to perform measurements because only “what gets measured gets done” as the saying goes in analytics. Few measurements are reported in the literature, however, typically based on standard statistical assumptions. In the case of a standard measurement of the bullwhip effect independence amongst the participating variables are required amongst others, but this is definitely known not to be perfectly true if any systematic control at all have taken place in the supply chain. This paper analyses the bullwhip measurement implications in case the standard test assumptions are violated and illustrates how to improve on the testing setup. This is further done with a special emphasis on the unavoidable small-sample aspects relating to such measurement in practice, which typically renders all statistical asymptotic or robust arguments quite unusable. It is shown how H_0-95%-confidence test intervals still easily can be obtained numerically in such cases.

1. Introduction

There has been considerable interest for some time over the many aspects of the bullwhip effect, its history, and its expected future (Geary et al., 2006) in relation to understanding the fundamental workings of a supply chain, at least on a theoretical level (Dejonckheere et al., 2003; Disney and Lambrecht, 2007). It is therefore only natural to proceed to the next stage to seek to measure this effect in practice. The understanding thus gained may then possibly be used for controlling and elimination purposes.

Many authors report bullwhip measures in order to put their theoretical work into perspective (Chen and Lee, accepted for publication, for an excellent review). Some have claimed that it may be an overestimated problem (Sucky, 2009). However, a few authors have actually made elaborate attempts to measure bullwhip effects empirically (Cachon et al., 2007; Fransoo and Wouters, 2000). Practical bullwhip measurement is an estimation process, which clearly involves statistical considerations. Any estimate is computed subject to certain assumptions that enable the assessment to be specific, however, relative to some overall uncertainty in measurement. Such measurement uncertainty typically stems from two sources: first, the lack of structural knowledge of what lies behind the measured phenomenon and, second, the general level of (white-)noise present in the applied data. Standard assumptions such as normality and independence are typically the basis of many standard tests, but are clearly only approximately satisfied in many situations.

A number of such measurement issues are relevant in many situations for the estimation of the bullwhip effect. Tests leading to a “significant” bullwhip effect measured based on the wrong premises are of course of no value. Here we consider small sample uncertainty issues in relation to bullwhip measurement. The issues and consequences are discussed and illustrated with a special focus on the H_0-95%-confidence test-interval and, as learning from this, thereby to provide a general recipe on how to deal with this measurement problem. The paper (Section 2) first considers generic issues in relation to bullwhip measurement in general, the effect of aggregation as well as where specifically in the chain the measurements are collected, followed by Section 3, where the focus is set on the general small sample problems in relation with estimating the bullwhip effect. Sections 4–6, respectively, deals with estimation issues when various types of stress and violation to the standard assumptions are at work—violation/stressing of the normality assumption, violation of independence in general terms and violation of independence formulated structural specifically. Section 7 then summarizes and concludes on the paper.

* Tel.: +45 8948 6322; fax: +45 8948 6660.
E-mail address: ehn@asb.dk

0925-5273/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpe.2012.08.004

2. Basic bullwhip measurement considerations

The measurement of a bullwhip effect may seem quite straightforward. If two sets of matching "in" and "out" time series data sets are available for demand and upstream orders, respectively, and these can be related in a meaningful way, the bullwhip effect (*BW*) can be defined, following the Chen, Drezner, Ryan and Simchi-Levi setup (Chen et al., 2000), simply as the ratio of two variances and can be computed as follows:

$$BW = \frac{Variance(Upstream\ Orders)}{Variance(Demand)} \tag{1}$$

It should, of course, be noted that the measure in (1) is not by far the only relevant performance metric in relation to measuring bullwhip phenomena (Beamon, 1998; Towill et al., 2007). However, it is definitely one of the most basic ones, at least since the paper of Chen, Drezner, Ryan and Simchi-Levi setup (Chen et al., 2000) was published.

The expression in (1) is clearly a measure that is specific to a transformational mechanism. It reveals if the transformational mechanism is dampening, neutral or amplifying in nature, i.e.

$$\begin{aligned} BW &< 1 \quad \text{implies signal dampening} \\ BW &= 1 \quad \text{implies signal neutrality} \\ BW &> 1 \quad \text{implies signal amplification} \end{aligned} \tag{2}$$

In general it is the amplification mode that is associated with a bullwhip situation. The transformational mechanism could be a specific single ordering mechanism but it could also be a system, which then represents a more macro-type transformation mechanism, e.g., a subsection of a supply chain covering a whole sequence of ordering schemes, where translation via bills of material as well as aggregation of streams of goods' flow can take place. Demand for a specific product or component can be an aggregate of several distinct demand sources. Also, the upstream orders for a specific component can be an aggregate of several distinct supply sources where several different ordering schemes are involved.

In such cases the *BW* measure is a conglomerate of amplification effects and the measure's ability, without further detailed knowledge, to provide insight for the purposes of control or elimination is most likely to be limited. An analogy is the checking of an individual's blood pressure. If the measurement is outside acceptable limits, further analysis is still required in order to pin-point the specific illness. The same issue arises somehow with such conglomerate *BW* measurement.

Part of the literature (Cachon et al., 2007) on actual *BW*-measurement problems focuses on the *BW* measurement performed with or without seasonal correction. The discussion here is whether the seasonal part of the demand can be taken care of by planning in advance or not. If it is possible to plan for the seasonal demand, this part of the total demand should of course not be included in the *BW* measurement. On the other hand if this distinction cannot be made, total demand is clearly relevant for obtaining a *BW* measure. Another issue relating to *BW* measurement that needs to be addressed is whether independent demand (specific variants of some finished product) or dependent demand (common product-components present in all variants) data (Vollmann et al., 2004) are being used for the analysis. Clearly if the bill-of-material (BOM) matrix between products and their required components is diagonal, then only a scaling in units takes place between independent and dependent demand and so the source of demand variability is unique. Scaling by the BOM translation in such cases is much like simple changes in measurement as for instance from tons to kilograms. Therefore, in order to make *BW* measurement independent of scaling or change in units etc., an extended *BW* definition should be applied in practice

$$V(X) = \frac{Variance(X)}{(Mean(X))^2} \tag{3}$$

and consequently

$$\begin{aligned} BW &= \frac{V(Upstream\ Orders)}{V(Demand)} \\ &= \frac{(Variance(Upstream\ Orders)/Mean(Upstream\ Orders)^2)}{(Variance(Demand)/Mean(Demand)^2)} \end{aligned} \tag{4}$$

In case the BOM matrix is not restricted to be a diagonal matrix, where more finished product variants demand sources mixes (BOM) into demand for the various product components, emphasis should primarily be on dependent demand in order to keep the focus on the upstream transformational aspects.

Having considered the measurement setup with respect to the product structure, there is still a question of the level of aggregation over products or product families, which also puts specific focus on a given ordering scheme. It defines whether a given *BW* measure might be suited for an overall assessment of the supply chains' dynamic functioning or "health", or might be more appropriate for monitoring how certain specific controls or ordering operations are functioning.

Also, proper attention must be given to the quantitative or statistical soundness of the actually obtained *BW* measurements. It must be kept in mind that any computation of a *BW* measure based on an observational fixed sample is simply an estimate of some underlying true *BW* measure. Any estimate has an error of measurement attached and must be interpreted accordingly. In a time series context, stationarity is a key concept. It states roughly that, being estimates, a computed variance and mean on a given sample have the intrinsic property that if the sample size were increased, the estimates would become more precise in relation to some unknown true underlying values (Box and Jenkins, 1976). This also implies that whenever data incorporate a trend, which is an element of strong non-stationarity, they have to be de-trended before a *BW* measurement is performed. Seasonality may or may not constitute a specific problem in this respect. However, if the incorporated seasonality turns out to have non-stationary characteristics, it clearly also has also to be removed. The detection of trend and seasonality related to singling out the non-stationary parts can often be somewhat tricky in practice, especially if we also include the concept of a stochastic trend, a pattern emanating for example from a pure "Random-Walk" process. Fortunately, a very simple and robust testing scheme exists – the Dickey–Fuller testing scheme (Dickey and Fuller, 1979) – which makes the task of securing stationarity in data by means of de-trending data-transformation of some sort, most likely differencing of some order, a manageable process.

3. The small sample measurement problem

As defined by formulas 3 and 4 above, the bullwhip measure appears to promise an easy and exact computation of what the bullwhip effect might be, given some observed "Upstream Orders" and corresponding "Demand" time series data samples. However, as noted above, this may be quite misleading as the computed value is simply an estimate of some underlying true parameter value, which is unobservable unless, of course, the sample size approaches infinity. Accordingly, what can be asked in a meaningful way is relative to some specified H_0-hypothesis, whether it is true or false given an observed data sample. For example, what could be asked is if the bullwhip effect is present or not in a statistical sense.