Int. J. Production Economics 177 (2016) 53–65

Contents lists available at ScienceDirect

Int. J. Production Economics

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

An empirically-simulated investigation of the impact of demand forecasting on the bullwhip effect: Evidence from U.S. auto industry

Chung-Yean Chiang [a,*], Winston T. Lin [b], Nallan C. Suresh [b]

[a] *Department of Logistics & Supply Chain Management, Georgia Southern University, P.O. Box 8036, Statesboro, GA 30460, United States*
[b] *Department of Operations Management and Strategy, University at Buffalo, SUNY, United States*

ARTICLE INFO

Article history:
Received 27 December 2014
Received in revised form
23 February 2016
Accepted 19 April 2016
Available online 20 April 2016

Keywords:
Empirical research
Bullwhip effect
Demand forecasting
Supply chain

ABSTRACT

This study empirically examines the impacts of three major aspects of demand forecasting on the magnitude of the bullwhip effect. Three research questions are addressed to investigate the association between (1) forecast accuracy, (2) aggregate forecasting, and (3) responsive forecasting and the bullwhip effect. Using forecasted demand generated from popular time-series forecasting models and real-life demand data, the study investigates the relationship between the forecasted results and the consequential bullwhip effect. The findings show that the forecasting methods used lead to the variation of the bullwhip effect. Moreover, the lead time reduction and the stable demand forecast are beneficial to reduce the bullwhip effect. However, our empirical results differ from previous findings in two ways: (i) improving forecast accuracy does not necessarily reduce the bullwhip effect and (ii) aggregate forecasting does not always reduce the bullwhip effect.

1. Introduction

The bullwhip effect (BWE) is described as the non-optimal solution adopted by supply chain participants to reach a global optimization for the whole system (Sterman, 1989) and defined as the increase of the variation of the order quantity from downstream members to upstream members in a supply chain (Lee et al., 1997b). Lee et al. (1997a) pointed out that the BWE is one major cause of the less-than-desired supply chain performance and has been observed in most manufacturing supply chains, such as Procter & Gamble's diaper products (Lee et al., 1997b), fast-moving consumer goods (Zotteri, 2013), and auto component in the assembly line (Klug, 2013). These cases among many others showed that the BWE is a long-lasting and still existing issue and can be observed in many different industries even after numerous efforts have been devoted to work on this problem for more than a decade. In addition, Lee et al. (1997a, b) showed that the BWE inflates supply chain operating cost by 12.5–25% and the elimination of the BWE would save the U.S. grocery industry for about $30 billion per year. In other words, the BWE is a serious problem across different industries and in different markets/countries. Summarizing the existing efforts, Geary et al. (2006) claimed ten major causes/principles of the BWE and demand forecasting is believed as a major contributor to the presence of the BWE (Barlas and Gunduz, 2011; Duc et al., 2008).

Forecasting-related factors were analytically/mathematically concluded to influence the magnitude of the BWE. With an assumed demand model, prior studies relied on a few forecasting methods to estimate the future demand which is further used to determine the inventory level and the order quantity for BWE estimation. The forecasting methods considered include simple moving average forecast (Chen et al., 2000b), simple exponential smoothing forecast (Chen et al., 2000a), double exponential smoothing forecast (Wright and Yuan, 2008), Holt-Winter's exponential smoothing forecast (Bayraktar et al., 2008), minimum mean-squared error forecast (Alwan et al., 2003; Zhang, 2004), autoregressive model (Chandra and Grabis, 2005), state-space approach (Gaalman and Disney, 2009), and damped trend exponential smoothing forecast (Li et al., 2014). Clearly, existing conclusions between forecasting and the BWE are drawn by a specific combination of the forecasting methods used and the demand pattern assumptions specified. As such, it would be interesting to investigate the forecasting-BWE problems by adopting new forecasting methods, by considering additional embedded forecasting properties, and by allowing different types of demand pattern.

The objective of this study is three-folds. The first is to provide a more insightful understanding between the BWE and selected aspects of the demand forecasting in response to the call for empirical verification of existing theories (Fisher, 2007) and offering managerial understandings between forecasting and operational

* Corresponding author.
E-mail addresses: cchiang@georgiasouthern.edu (C.-Y. Chiang), mgtfewtl@buffalo.edu (W.T. Lin), ncsuresh@buffalo.edu (N.C. Suresh).

http://dx.doi.org/10.1016/j.ijpe.2016.04.015
0925-5273/Published by Elsevier Ltd.

performance (Fildes and Kingsman, 2010). Prior studies tested one or two aspects mentioned above at a time without taking a comprehensive view to demand forecasting. For example, some studies suggested a negative relationship between forecast accuracy and the BWE (Alwan et al., 2003; Bayraktar et al., 2008; Zhang, 2004). In other words, it showed that the smaller the forecast error, the smaller the BWE will be. Others (Barlas and Gunduz, 2011; Bayraktar et al., 2008; Chen et al., 2000a) have suggested that the responsiveness of the forecast which is influenced by the smoothing parameter is positively related to the strength of the BWE. Again, the conclusions are relied on an imposed demand assumption which may limit the reliability of the application in the real-world. Moreover, existing forecasting-oriented studies suggested that the relationship between forecast accuracy and the forecast utility is complex, especially in an inventory management setting which affects the order volume and inventory level at hand (Ali et al., 2012; Syntetos et al., 2010). These studies empirically explored the association between the forecast accuracy and the cost but little was made to study the relationship between the forecast accuracy and the BWE. Hence, we use the real-world demand to empirically verify the previous findings in a more comprehensive picture. Thus, this study provides additional findings to the accuracy-utility relationship.

The second objective is to investigate whether the aggregate forecasting leads to less BWE for the company. Current studies emphasized on testing whether or not the use of aggregate data to measure the BWE leads to stronger BWE. For example, Cachon et al. (2007) claimed that “whether data aggregation may preserve or mask the BWE is dependent on the correlation of the production and demand across the units being aggregate (firms, products, etc.)”. Likewise, Chen and Lee (2012) argued that aggregating products with similar seasonality leads to less BWE. However, to the authors' best knowledge, no efforts were made to investigate the impact of aggregate forecasting on the BWE. In other words, it leaves an interesting question that remains to be answered, namely, can the BWE be managed by using data-aggregated forecasted demand. This study intends to fill this gap by considering demand at multiple levels and measuring the BWE from various groups.

The third objective is to re-examine the impact of lead time on the BWE when a new type of demand process is considered. The BWE was proposed to be a result of a joint effect from non-zero lead time and demand forecast (Forrester, 1958). Many existing efforts suggested that, regardless of the forecasting policy used, the longer the lead time, the stronger the magnitude of the BWE (Chen et al., 2000b; Zhang, 2004). However, Duc et al. (2008) have suggested “the magnitude of the BWE does not permanently increase when the lead time increases and claimed that the positive association between the length of the lead time and the magnitude of the BWE is greatly dependent on the attributes of the considered demand process”. It implies that the claimed relationship between the length of lead time and the BWE should be retested when a new demand process is considered. Hence, since this study considers a less controlled demand processes presenting the complexity of the real world demand, it is interesting to re-evaluate the relationship between the length of the lead time and the magnitude of the BWE.

The remainder of the paper is organized as follows. Section 2 presents the research methods and models. In this section, we briefly introduce the forecasting methods, the replenishment policy, and the bullwhip measure. In Section 3, a description of the data and data source is provided. Section 4 presents the empirical results and describes the implications. Finally, the last section offers concluding remarks and extensions for future research.

2. Research methodology

In this section, we depict the business settings, provide the information concerning the inventory policy used and the estimation the order quantity, present the forecasting policies considered and other tested factors, and describe how to investigate the impact of aggregate forecasting upon the BWE.

2.1. The business settings and the BWE measure

Due to data availability, we conduct this study by emphasizing on the operations at the retailer (auto dealer) echelon in the automotive supply chain and use simulation to estimate the needed information. It is assumed that the retailer satisfies the consumer demand from its own on-hand inventory and refills the inventory by placing orders to the upstream party which is usually the auto maker. The upstream party is assumed to have no capacity restraint and is able to satisfy all order quantity from the retailer. Finally, it is also assumed that there is no negative order quantity placed by the retailer, meaning that the retailer cannot return the inventory to the manufacturer.

The timing of events in a period is described as follows. At the beginning of the period, the retailer receives the order placed l periods ago. The retailer satisfies the end customer's demand which is present as the actual monthly sales by the available inventory. The forecasted demand for the next period is generated based upon the actual sales information at the current period. Given the forecasted demand and the chosen inventory policy, the retailer places the order which will be received at the beginning of l periods after, where l is the lead time. To be clear, it is assumed that lead time is zero when l is one as used in Chen et al. (2000b). For example, if a retailer which places its order at the end of January receives this January order at the beginning of February, it is considered to have zero lead time. Yet, since there is one time period difference between the time the order is placed and the time the order is received, we consider that the order is delivered one time period late ($l=1$). Given the time frame for each move, we will use historical monthly demand at the auto type level to forecast future demand. Then, based on the forecasted demand, we can estimate the order quantity and the forecast error. In the end, when all forecasted demand and order quantity in 2008 are generated, the 2008 BWE is calculated.

The BWE measure is used to evaluate the influences of demand forecasting and replenishment lead time on the supply chain performance. In this work, the BWE at time t is evaluated using Lee et al. (1997b)'s bullwhip measurement which is the ratio of the variance of order (O_t) placed by the retailer at time t to the variance of demand (D_t) satisfied by the retailer at time t as described by

$$\textit{Bullwhip effect}_t = var(O_t)/var(D_t), \tag{1}$$

where D_t is the monthly sales obtained from data while O_t is the order quantity estimated from the inventory policy considered.

2.2. The replenishment policy

In light of the popular application in previous BWE studies and the nature of the data used in this study, a periodic order-up-to (OUT) method is used to estimate the order quantity. Although the OUT policy is a likely cause to the presence of the BWE, the OUT policy is the standard ordering mechanism in many MRP systems (Gilbert, 2005), and is not restricted to specific demand types and can be used with the existence of trend (Li et al., 2014). Besides, the objective of this study is not to examine the existence of the BWE. Hence, the OUT policy is treated as the replenishment policy.

To implement the OUT policy, the management periodically reviews the inventory position and places an “order” to recover the inventory position “up-to” a pre-determined level. The order quantity under the OUT replenishment policy, as considered in Chen et al. (2000b), is shown as follows:

$$O_t = y_t - y_{t-1} + D_{t-1}, \tag{2}$$

where y_t is the order-up-to point and is estimated from the observed demand as

$$y_t = \hat{D}_t^{\,l} + z\hat{\sigma}_{et}^{l}, \tag{3}$$

$\hat{D}_t^{\,l}$ is an estimate of the lead time demand which is calculated by multiplying the number of lead time (l) and the forecasted demand in the next period, $\hat{\sigma}_{et}^{l}$ is an estimate of the standard deviation of the L period forecast error, and z is a constant chosen to meet a desired service level which is assumed at 95% for this study. We used the past 12-period forecast errors to estimate the standard deviation of the forecast error since the auto maker offers new models annually. The service level is arbitrarily chosen and does not make significant influence since the same service level applies to all tested auto types and the forecasting methods used to estimate the order-up-to position. As a result, the order quantity in each period and the BWE in 2008 can be estimated.

2.3. The demand forecasting methods and forecast accuracy

The impact of demand forecasting is the major concern in this study. Forecasted demand serves as the major trigger for most supply chains to develop the replenishment decisions. Inaccurate forecasts lead to over- or under-stock in the supply chain which further shows a lagged effect on the future replenishment decisions of the focal firm and the upper-level supply chain partners. This study uses SAS 9.3 to estimate all forecasted values under various forecasting methods considered.

2.3.1. Forecasting methods

McCarthy et al. (2006) showed that firms had less-than-expected forecast performance because they relied on using simple forecasting methods, such as moving average, simple exponential smoothing, and straight-line projection, though more advanced methods, like regression-based forecasting, times series forecasting, and simulation, are known and accessible. Likewise, analytical supply chain studies testing the impact of forecasting policy on the BWE merely considered simple quantitative forecasting techniques (Chandra and Grabis, 2005; Wright and Yuan, 2008; Zhang, 2004) and concluded that demand forecasting has impacts on the BWE. Thus, this study, drawing from prior findings and drawbacks, proposes to empirically test the influence of various widely used forecasting methods upon the development of the BWE. The select forecasting techniques include the Naïve method, simple moving average, simple exponential smoothing, double exponential smoothing Brown's (DES-Brown) and Holt's (DES-Holt), seasonal exponential smoothing, damped trend exponential smoothing (damped ES), Holt-Winters multiplicative exponential smoothing, and univariate multiplicative seasonal autoregressive integrated moving average (SARIMA) model. Among the methods mentioned except for SARIMA model, damped ES has often been promoted as the most accurate forecasting technique in the M-competitions (Li et al., 2014; Makridakis and Hibon, 2000). In fact, using transfer function on damped forecasting and the OUT policy, Li et al. (2014) concluded that it can eliminate the BWE when demand is a single harmonic frequency, given the value of smoothing parameters (level, trend, and damping) fall in the bullwhip avoidance regions. We, hence, consider that the BWE from damped ES is a good benchmark for comparison to tell whether there is additional value to use advanced forecasting technique to manage the BWE. Since most of the tested forecasting policies are well-known and popularly used, we merely provide the concept of SARIMA in this section. For the other considered forecasting models, the detailed information is outlined in the Appendix A.

Box et al. (1994) suggested that monthly sales data are often characterized by seasonality, especially in the automotive industry (Waller, 2004). Due to the possible sales trend and/or seasonality, it is critical to first recognize whether the considered time series variable is stationary or non-stationary by conducting the unit root tests and observing the historical sales diagrams. If the data employed present strong seasonality, we then follow the Box-Jenkins approach to estimate the SARIMA model; and the demand process is adequately and reasonably present in the multiplicative seasonal form as described in Eq. (4) below:

$$\phi_p(B)\Phi_P\left(B^S\right)\nabla^d\nabla_S^D D_{i,t} = \theta_q(B)\Theta_Q\left(B^S\right)\varepsilon_{i,t}, \tag{4}$$

where

$$\phi_p(B) = \left(1 - \phi_1 B^1 - \phi_2 B^2 - \ldots - \phi_p B^p\right),$$

$$\theta_q(B) = \left(1 - \theta_1 B^1 - \theta_2 B^2 - \ldots - \theta_q B^q\right),$$

$D_{i,t}$ is the unit sales for auto type i at time t, B is the backward shift operator, ϕ_i and Φ_i are respectively the autoregressive parameters for regular and seasonal time series model for auto type i, θ_i and Θ_i are respectively the moving average parameters for regular and seasonal model for auto type i, $\varepsilon_{i,t}$ represents the error term for auto type i in the time series SARIMA model, and S is the number of seasons in a fixed time period. The errors are generally assumed to be independently, identically, and normally distributed, satisfying three required conditions: zero mean, constant variance, and no covariance between any two errors. Such errors are known as white noise.

Prior studies have concluded that forecasting methods used have impacts on the BWE. This study examines this conclusion through an ANOVA analysis to compare the average BWE across all tested forecasting methods.

2.3.2. Forecasting accuracy and BWE

As mentioned above, conventional wisdom suggests that the forecast accuracy has a negative relationship with the BWE. This study tests this relationship in two ways: (i) a rank analysis to reveal the ranking relationship and (ii) a Pearson correlation test to measure the numerical association between the BWE and the forecast accuracy among all considered auto types. Previous studies (e.g., Zhang, 2004) showed that if a minimum-mean-squared-error (MMSE) forecast is made to make replenishment decisions under OUT policy, the BWE is minimized. Hence, the forecast method having the minimum forecast error would rank highest and should lead to the least BWE. The Pearson correlation test is undertaken to assess the likely association between the accuracy and the BWE. The mean absolute percentage error (MAPE) is used to determine which forecasting policy produces the most accurate forecast. MAPE is a popularly-applied measure to evaluate the difference between the estimated value and the actual value. Moreover, MAPE can be used to compare the forecast accuracy across different auto types. The formula of MAPE is given by

$$\text{MAPE} = \sum_{t=1}^{n} \frac{\left|D_t - \hat{D}_t\right|}{D_t} / n \tag{5}$$

where $\hat{D}_t$ is the forecasted demand while D_t is the actual demand.

2.3.3. Other forecast-related issues and BWE

Responsive or stable forecast, in addition to the forecast accuracy, is another major property when predicting future demand and making replenishment plans. Using responsive forecast makes firms able to hold less safety stock and set a low order-up-to quantity since the forecast responds to the demand change quickly. The choice of responsive forecast can be made by reducing the number of periods used when moving average is used or by increasing the value of the smoothing parameter when the simple exponential smoothing is applied. In order to investigate the impact of responsive and stable forecasts upon the BWE, this study uses a sensitivity analysis to measure the BWE when three selected smoothing parameters are applied to forecast the future demand.

Lead time is another long-believed factor having influences on the BWE. However, the existence of such a relationship between lead time and BWE was challenged by Duc et al. (2008). Because of the inconsistent conclusions regarding the impacts of the lead time on the BWE, this study also uses sensitivity analysis to test the lead time-BWE relationship by introducing the lead time effect into the decision process of the order-up-to quantity.

2.4. The impacts of aggregate forecasting

The level of aggregation is one of the two critical factors needed to be determined before making a forecast (Krajewski et al., 2012). Forecasting from aggregate data is, in general, more accurate than that from disaggregate data since aggregation helps to remove the individual-specific variation. Current studies claimed that time aggregation and product aggregation help to reduce the BWE when using the secondary data to directly measure the BWE (Cachon et al., 2007; Chen and Lee, 2012). This study, instead of testing the change of the BWE under different aggregation levels, aims to uncover the effect of aggregate forecasting on the BWE. This study considered three levels for an auto brand: type level, group level, and brand level. From the type level to group level, we combined the monthly sales of sedans and sports car to form the sales of the "Car" group and the sales of SUVs, Pickup trucks, Van, and Wagon to create the sales of the "Truck" group. Then, from group level to brand level, the sales of "Light vehicle" is developed by combining the sales of "Car" group and "Truck" group. A graphical hierarchy is provided below to show the levels considered in this study (Fig. 1).

Next, the same forecasting methods used to make forecast on the type level were applied on the "Car", "Truck", and "Light vehicle" to estimate the forecasted monthly demand, forecast accuracy, and the BWE in 2008. Finally, the Fisher's LSD test and the Tukey's HSD test were used to evaluate whether the BWEs from the aggregate data are significantly different from those obtained from using the disaggregate data.

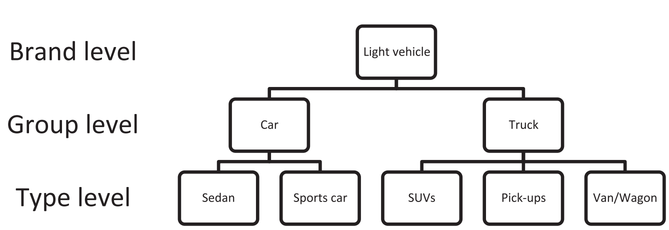

Fig. 1. A hierarchy of considered auto level.

3. Data description and summary

This section provides the information about the data used. In the data description subsection, the data source is provided. In the data summary subsection, a table is used to describe the demand characteristics of the selected auto types.

3.1. Data description

The data used for this research were obtained from Automotive News: Annual Market Data Book. We carefully reviewed the information of U.S. monthly unit sales (1991–2008) and decided to use the data to present the demand for the selected Chevrolet auto types. In other words, we considered one demand time-series for each considered auto type. There are two reasons why we only consider the auto models from Chevrolet. First, among all auto brands under General Motors (GM), Chevrolet had the highest sales volume in 2008 (5.2% for Chevrolet; 2.7% for GMC; 2% for Pontiac; 1.4% for Saturn; 1.2% for Cadillac; 1% for Buick for the 2008 auto market) and is the continuous brand after the reorganization in 2009. Second, among the major US-based automakers, GM took the most market share in the U. S. light vehicle market. The firm had the second largest market share (13.5%; 13.9% for Toyota group) of light vehicle sales in 2008. With the brand's long history and market position, we believe that vehicles from Chevrolet are good candidates to conduct this study.

Other than the reasons to choose a sufficiently large number of observations in order to cover a complete economic cycle and generate reliable results, the authors offer two more reasons for choosing the 1991–2008 sales data to measure the demand. First, the company filed a bankruptcy protection and experienced a reorganization in 2009. The reorganization changed the firm's structure significantly and created substantial external influences on its demand. Moreover, the recent recession hurt the economy and changed the demand pattern enormously. This sudden change in the market might lead to unexpected impacts on forecasts and, hence, on the forecasting-BWE relationship. Therefore, we decided to conduct this study based on the data for the time period from 1991 to 2008.

3.2. Data summary

As mentioned in Cachon and Olivares (2010), auto models can be classified into certain types according to the product attributes. In order to improve the reliability of the test results, we selected ten auto types in the light of vehicle categories as outlined in Table 1. The economy sedan and the pony car are ignored because of

Table 1
The summary of selected auto types.

Type	Demand uncertainty 1991–2008	Demand uncertainty 2008	Auto model example
Compact sedan	0.2865	0.3434	Cobalt
Mid-size sedan	0.3134	0.1860	Malibu
Full-size sedan	0.2702	0.2556	Impala
Sports car	0.3431	0.3875	Corvette
Mid-size SUV	0.3762	0.4285	Trailblazer
Full-size SUV	0.5967	0.2644	Tahoe
Mid-size pick-up truck	0.3926	0.2572	Colorado
Full-size pick-up truck	0.2336	0.1968	Silverado
Van	0.2222	0.2166	Van
Wagon	0.4069	0.2052	Suburban
Car	0.1844	0.2023	
Truck	0.2443	0.1770	
Light vehicle	0.1982	0.1704	

a short-term shut-down during the period considered. The demand uncertainty for each tested auto type is represented by the coefficient of variation equal to the ratio of the standard deviation of the sales to the mean of the sales. It measures the relative dispersion and compares the attributes among multiple time series. As expected, the sports car presents the highest level of demand uncertainty among cars. For the trucks, the full-size SUV shows the highest demand uncertainty in the 1991–2008 period while the coefficient of variation for the mid-size SUV is the highest in year 2008. In general, the SUV family, including the SUVs and the wagon, has greater demand variability than the pickup trucks and the van. Among the ten types of the light vehicles considered, the full-size sedan and the van demonstrate similar demand uncertainty for the two time periods. Except for the compact sedan, the sports car, and the mid-size SUV, the rest of auto types show larger demand uncertainty in the 1991–2008 period than in the year 2008. At the group level, Car has the lower demand uncertainty during the 1991–2008 period while Truck has the opposite observation. Finally, when considered at brand level, the demand uncertainty (0.1982) in the 1991–2008 period is also larger than that (0.1704) in the year 2008.

4. Results and discussion

In this section, firstly we present the critical information concerning the forecasting methods considered. Secondly, the BWE estimates are provided for all auto types. Thirdly, the association between forecasting methods and the BWE is discussed. Fourthly, the results of sensitivity analysis are reported to demonstrate the impacts of lead time and forecasting responsiveness on the BWE. Fifthly and finally, the results from the Fisher's LSD and Tukey's HSD tests are presented to examine the influence of aggregate forecast on the BWE.

4.1. SARIMA models

The application of the SARIMA is new to the BWE field. Although it is the first time that SARIMA is used to study BWE, we simply summarize the key test results and validate the estimation outcomes because the major objective is to investigate the properties of this new model, instead of explaining how to use it.

When using SARIMA, it is necessary to check the stationarity of the monthly sales data of each chosen vehicle prior to moving to the time series estimation process. To test whether the time series variable is stationary, two methods are often employed: the time plot of the variable and the unit root test. If the time plot reveals the patterns of trend and seasonality, one can conclude that the time series variable under consideration is non-stationary. Four charts are provided in Fig. 2 to shows the observed four major patterns among ten auto types. These patterns include hilltop-down change (compact sedan), horizontal movement (mid-size sedan), upward trend (sports car), and inverted U-shape (mid-size SUV). From the sales charts, due to the observable sales trend and seasonality, the demand process of these selected auto types is tentatively considered to be non-stationary. Fig. 2 depicts the time plots of unit sales for four selected auto types from 1991 to 2008.

In addition to the time plot, the unit root test is the other widely-applied tool to determine whether or not a time-series variable is stationary. SAS 9.3 was used to test the stationarity, estimate the SARIMA model, and check for the residual requirement. Two unit root tests used in this study are the augmented Dickey-Fuller test and the Phillips–Perron test, because of their availability of the software programs and their recognition as the major detection tool (Box et al., 1994). The test results confirm that the original data of sales are non-stationary, thereby supporting our previous claims of the existence of the trend or the seasonality in sales. Since the seasonality effect on sales is strong in the automotive industry (Waller, 2004), we then tested the existence of unit root for the 12-month seasonally differenced sales data and found that the unit root does disappear. Therefore, in the identification stage, it is concluded that the demands (sales) for all ten

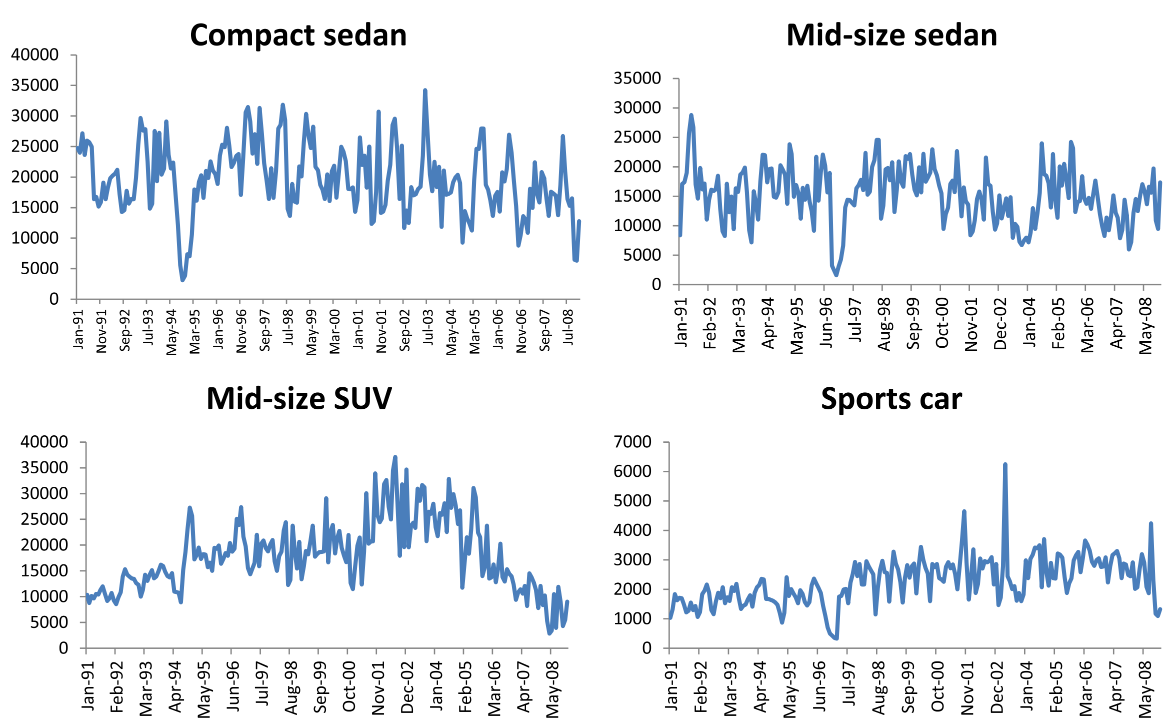

Fig. 2. The monthly sales information for selected Chevrolet auto types.

Table 2a
The magnitude of bullwhip effect without lead time at the type level.

Forecasting model	Compact sedan		Mid-size sedan		Full-size sedan		Sports car		Van		Wagon	
	BWE	Rank	BWE	Rank	BWE	Rank	BWE	Rank	BWE	Rank	BWE	Rank
Naïve	2.51	9	3.95	10	5.01	10	4.50	10	5.53	9	5.56	9
MA(3)	1.76	3	1.90	4	2.43	2	1.99	8	2.36	5	1.73	7
MA(6)	1.32	1	1.57	1	1.31	1	1.46	3	1.74	1	1.47	6
Simple ES	2.31	6	2.42	6	3.17	6	1.64	6	2.85	8	1.41	5
DES_Brown	2.57	10	1.86	3	2.73	3	1.19	1	1.76	2	1.37	3
DES_Holt	2.31	6	2.42	6	3.17	6	1.53	4	2.84	6	1.39	4
Seasonal ES	1.92	4	2.61	9	3.29	9	2.05	9	2.02	3	1.76	8
Damped ES	2.31	6	2.42	6	3.17	6	1.53	4	2.84	6	1.36	2
Holt-Winters	1.73	2	2.35	5	3.11	5	1.35	2	2.17	4	1.08	1
SARIMA	1.92	4	1.85	2	2.88	4	1.92	7	6.23	10	6.23	10
Average	2.07		2.33		3.03		1.92		3.03		2.33	
Forecasting Model	Mid-size SUV		Full-size SUV		Mid-size Pickup truck		Full-size pickup truck		Summary Statistic			
	BWE	Rank	BWE	Rank	BWE	Rank	BWE	Rank	Rank	Min	Max	Range
Naïve	5.88	10	4.43	10	1.71	7	5.79	10	10	1.71	5.88	4.16
MA(3)	2.31	8	2.02	8	1.49	2	2.55	9	6	1.49	2.55	1.06
MA(6)	1.41	1	1.36	3	1.01	1	1.70	7	1	1.01	1.74	0.73
Simple ES	2.02	4	1.54	7	1.62	4	1.37	5	8	1.37	3.17	1.80
DES_Brown	1.88	3	1.50	6	1.95	8	1.43	6	4	1.19	2.73	1.54
DES_Holt	2.02	4	1.49	5	1.65	6	1.30	2	5	1.30	3.17	1.87
Seasonal ES	2.24	7	1.25	1	3.11	9	1.32	4	8	1.25	3.29	2.05
Damped ES	2.02	4	1.48	4	1.62	4	1.30	2	3	1.30	3.17	1.87
Holt-Winters	1.76	2	1.27	2	1.50	3	1.20	1	2	1.08	3.11	2.03
SARIMA	2.69	9	3.19	9	3.55	10	2.12	8	9	1.85	6.23	4.38
Average	2.42		1.95		1.92		2.01					

auto types are non-stationary and seasonal difference is required to identify a fitted model in the estimation stage.

At the identification stage, a few possible SARIMA models may be identified. Among those possible models, Akaike information criterion and Schwarz information criterion were applied to select a final candidate. As a result, thirteen estimated SARIMA models are all checked for the minimized criterion value and white noise requirements upon the residual term, meaning that the error term is tested to be normally distributed and has no autocorrelation. The details are offered in Appendix B.

4.2. BWE estimation

The objective of this study is to investigate the relationship between the BWE and forecasting. To estimate the BWE for each auto type in 2008, we used the "Time Series Forecasting System" function in SAS 9.3 to estimate monthly demands and MAPEs for all considered forecasting techniques. One advantage for using this function is that the program can automatically identify the optimal value for each smoothing parameter to minimize the forecast error and, simultaneously, we are permitted to arbitrarily specify the values of the parameters to carry out sensitivity analysis. After the demand was generated, the order quantity and the BWE were calculated. The BWE measures when the lead time does not exist are reported in Tables 2a and 2b below.

Table 2b
The magnitude of bullwhip effect without lead time at the group level and brand level.

Forecasting model	Car		Truck		Light vehicle	
	BWE	Rank	BWE	Rank	BWE	Rank
Naïve	3.00	10	6.70	10	4.53	10
MA(3)	1.68	6	2.56	8	1.90	9
MA(6)	1.33	3	1.79	7	1.32	6
Simple ES	2.22	7	1.65	4	1.17	3
DES_Brown	1.00	1	1.71	5	1.21	5
DES_Holt	2.22	7	1.43	3	1.05	2
Seasonal ES	1.51	5	1.71	5	1.53	7
Damped ES	2.22	7	1.40	2	1.02	1
Holt-Winters	1.41	4	1.39	1	1.19	4
SARIMA	1.22	2	2.60	9	1.79	8
Average	1.78		2.29		1.67	

A few findings can be recognized from Tables 2a and 2b. First, the numerical BWE-value changes among every auto type and forecasting-method pair, and the range of BWE values lies between 1.01 and 6.23 across all ten types. This evidence suggests that the BWE always exists at the product level in the retailer echelon. This finding is similar to Dejonckheere et al. (2003)'s saying that as long as the OUT policy is used, the BWE always exists due to the necessary safety stock held. In contrast, the finding is against the conclusion of Li et al. (2014) that the BWE can be avoided by using damped ES jointly with the OUT policy.

Second, these BWE values not only show the absolute volume of increasing order-to-demand variation but also the relative strength between any two type-method pairs. For example, when MA(6) is used to predict the demand for the compact sedan, the variation of order quantity is 132% more than the variation of demand quantity. Meanwhile, comparing the BWE measures of MA(6) and Naïve for the compact sedan, we find that the BWE from Naïve is almost twice as strong as that from MA(6).

Third, by taking a rank analysis at the type level, we found that using MA(6) may obtain the smallest BWE while the Naïve forecast shows the largest (refer to the overall rank column). Among ten considered forecasting methods, the Naïve method consistently generates an unfavorable BWE performance which is greater than 2 for nine out of ten types. In addition, SARIMA shows less-than-expected BWE performance, especially among all truck-based types, and ranks the second worst. On the contrary, MA

(6) presents some best BWE performance and, generally speaking, acceptable BWE performance when the rank is low. Finally, although prior studies attested that damped ES is the most accurate method and is a useful tool to control BWE, MA(6) and Holt-Winters, surprisingly, are found to outperform damped ES.

Fourth, the ranges of BWE differ among ten tested forecast methods. MA(6) has the narrowest range while SARIMA has the greatest. The finding implies that although the BWE cannot be diminished, it is possible to control the variation of BWE across different product line in a small range if the right forecasting method is applied.

Fifth, although seasonality exists in the automotive market, we are surprised to observe that the tested forecasting methods which consider the seasonality, namely, seasonal exponential smoothing, Holt-Winters, and SARIMA, do not necessarily outperform those which do not.

Sixth, the above-mentioned findings at the type level do not necessarily apply to the group level and brand level. For example, although Naïve and SARIMA still show unfavorable results, MA (6) does not perform as the best model, indicating that the level of aggregation has influences on managing BWE.

Seventh, looking at the BWE from the product perspective, we found that the sports car and the mid-size pickup have the lowest average BWE, while the full-size sedan and the van have the highest. Using the group level results reported in Table 2b, "Car" shows less BWE than "Truck" does in 2008. This finding is interesting since the 2008 demand uncertainty of "Car" is greater than that of "Truck" and it deems to provide us with a plausible conclusion of the negative relationship between the demand uncertainty and the BWE.

From the findings in Tables 2a and 2b, it is unexpectedly found out that although SARIMA is an advanced and powerful method which involves trend and seasonality, it does not help to control the BWE, especially for trucks. The test results in this study seem to get derailed from conventional wisdom that using an advanced forecasting method leads to better forecast accuracy (McCarthy et al., 2006) which in turn leads to less BWE (Duc et al., 2008; Jaipuria and Mahapatra, 2014). Thus, we decide to further re-validate the previous claim by investigating the relationship between forecast accuracy and the BWE.

4.3. The influence of forecasting methods upon BWE

Here, we report the results showing the influence of using different forecasting techniques upon the BWE. The ANOVA results presented in Table 3 confirm the relationship between the BWE and the forecasting methods as claimed by Zhang (2004). To be

Table 3
ANOVA results for the causes of the BWE.

Source	DF	Sum of squares	Mean square	*F* value	*Pr* > *F*
BWE-method					
Model	10	88.3265789	8.8326579	13.28	< 0.0001
Error	99	65.8674089	0.6653274		
Total	109	154.1939874			
R-square	Root MSE	BWE mean			
0.572828	0.815676	2.183294			
BWE-type					
Model	9	14.8511264	1.6501252	1.18	0.3134
Error	100	139.342861	1.3934286		
Total	109	154.1939874			
R-square	Root MSE	BWE mean			
0.096315	1.180436	2.183294			

specific, 57.28% of the variation of the BWE can be explained by the use of different forecasting methods. However, virtually no relationship between the BWE and the auto type is formed if 9.63% of the variation of the BWE explained by auto types is considered negligible or trivial.

Given the test conclusion from Table 3, we are stimulated to further investigate the question as to which factor is the major cause for the strong association between the BWE and the forecasting methods. The first factor to be tested is the forecast accuracy measured by MAPE. In Tables 4a and 4b, MAPEs and the ranks based on the MAPE are provided to assess the forecastability of each forecasting technique. The lower the value of MAPE, the less the forecasting error is expected and the better is for the rank of a forecasting technique. In other words, the forecasting policy with the lowest MAPE is believed to generate the forecasts which are considered the closest to the actual values. The forecastability of each forecasting technique is analyzed individually as a type, comprehensively across all ten considered auto types, and aggregately at three different levels. First, the SARIMA generates accurate forecasts for most car types but poor forecasts for truck types. Next, based on the rank analysis, the Naïve method is the best while MA(6) and DES_Holt share the worst, for all ten auto types considered. Third, the average MAPE for each auto type represents an overall picture depicting whether or not one forecasting method is capable of offering accurate forecasts to an auto type. The mid-size sedan has the lowest average MAPE while the mid-size SUV has the highest. Fourth, looking at the group level and brand level, we observe that SARIMA provides the best forecast for Car, MA(3) produces the best forecast for Truck, and MA(6) generates the best forecast for Light vehicle. Fifth, the average MAPE is smaller for Light vehicle than that for Car and Truck. Meanwhile, the average MAPE for Car is smaller than that of the four car types considered and the average MAPE for Truck is smaller than five of the six considered truck types.

Previous studies argued that using a MMSE-forecast helps to control the BWE (Zhang, 2004) and that the better the forecast, the less the BWE (Duc et al., 2008). Table 5 offers the information about the ranking of the demand uncertainty, the forecast accuracy, and the average BWE for each auto type in 2008. In terms of the ranks between the average BWE and the forecast accuracy, our test results do not necessarily support the existing wisdom. Then, we conduct the Pearson correlation analysis and provide the results in Table 6a and Table 6b to verify the claimed accuracy-BWE association. Table 6a gives the Pearson correlation results when all BWE measures and MAPEs are considered at the same time, regardless of the auto types and the forecast methods applied. To control the possible effect of the non-optimized SARIMA forecast, we conducted the correlation test in two scenarios: (i) SARIMA included (10 auto types by 10 forecast methods) and (ii) SARIMA excluded (10 auto types by 9 forecast methods). As observed from Table 6a, the correlation is not significantly different from zero for both scenarios. Then, the correlation between BWE and forecast accuracy for each individual auto type was tested as shown in Table 6b. Our results show that four of the ten auto types (sports car, full-size SUV, and full-size pickup truck are the common three) earn significant correlations no matter whether SARIMA forecast is included. Moreover, in Table 6a, no matter whether SARIMA is included, the sign of the correlation coefficient is not positive and, hence, is contrary to the existing understanding that the greater the forecast accuracy (meaning smaller MAPE), the less the BWE. Moreover, the test results are even more interesting and unexpected in that seven of the eight significant correlations in Table 6b are negative. In other words, our results suggest that the less accurate the forecast (meaning larger MAPEs), the less the BWE. This finding is consistent to Hosoda and Disney (2009)'s conclusion that poor forecast accuracy is not always bad for the

Table 4a
MAPEs at the type level in year 2008.

Forecasting Model	Compact sedan		Mid-size sedan		Full-size sedan		Sports car		Van		Wagon	
	MAPE	Rank	MAPE	Rank	MAPE	Rank	MAPE	Rank	MAPE	Rank	MAPE	Rank
Naïve	0.3013	3	0.2271	7	0.2790	5	0.3402	1	0.3147	7	0.3168	5
MA(3)	0.3753	9	0.2343	8	0.3346	10	0.4509	5	0.2489	1	0.2734	2
MA(6)	0.4296	10	0.2456	9	0.2656	3	0.4539	6	0.2697	6	0.3317	8
Simple ES	0.3288	6	0.2186	4	0.2930	6	0.4437	4	0.2599	2	0.3364	9
DES_Brown	0.3418	8	0.2460	10	0.2986	9	0.4813	10	0.2659	5	0.3056	4
DES_Holt	0.3280	5	0.2187	6	0.2939	8	0.4594	8	0.2620	4	0.3509	10
Seasonal ES	0.2949	2	0.2003	3	0.2592	2	0.3632	3	0.3158	8	0.2606	1
Damped ES	0.3288	6	0.2186	4	0.2931	7	0.4568	7	0.2614	3	0.3280	7
Holt-Winters	0.3051	4	0.1969	2	0.2756	4	0.4789	9	0.3167	9	0.3024	3
SARIMA	0.2873	1	0.1343	1	0.2201	1	0.3629	2	0.3568	10	0.3218	6
Average	0.3321		0.2140		0.2813		0.4364		0.2872		0.3127	

Forecasting Model	Mid-size SUV		Full-size SUV		Mid-size Pickup truck		Full-size pickup truck		Total
	MAPE	Rank	MAPE	Rank	MAPE	Rank	MAPE	Rank	Rank
Naïve	0.6323	3	0.3078	1	0.1608	1	0.1884	1	1
MA(3)	0.5976	2	0.3243	2	0.2475	7	0.2103	2	3
MA(6)	0.6582	9	0.3869	7	0.2763	9	0.2229	3	9
Simple ES	0.6365	5	0.3831	6	0.1747	3	0.3060	6	4
DES_Brown	0.5450	1	0.3478	4	0.1933	6	0.2807	4	7
DES_Holt	0.6446	8	0.4283	9	0.1734	2	0.3474	10	9
Seasonal ES	0.6391	7	0.3416	3	0.3475	10	0.2924	5	2
Damped ES	0.6348	4	0.4200	8	0.1750	4	0.3262	7	6
Holt-Winters	0.6375	6	0.4393	10	0.1931	5	0.3335	9	7
SARIMA	0.7270	10	0.3661	5	0.2677	8	0.3275	8	5
Average	0.6353		0.3745		0.2209		0.2835		

Table 4b
MAPEs at group and brand level in year 2008.

Forecasting Model	Car		Truck		Light vehicle	
	MAPE	Rank	MAPE	Rank	MAPE	Rank
Naïve	0.1988	4	0.2160	2	0.1769	4
MA(3)	0.2434	10	0.2015	1	0.1703	2
MA(6)	0.2191	6	0.2291	4	0.1673	1
Simple ES	0.2260	7	0.2946	9	0.2300	10
DES_Brown	0.2032	5	0.2357	6	0.1924	7
DES_Holt	0.2260	7	0.2162	3	0.1739	3
Seasonal ES	0.1737	2	0.2878	7	0.2202	9
Damped ES	0.2260	7	0.2323	5	0.1911	6
Holt-Winters	0.1743	3	0.3085	10	0.1802	5
SARIMA	0.1408	1	0.2930	8	0.1986	8
Average	0.2031		0.2515		0.1901	

Table 5
Summarized ranks for demand uncertainty, forecast accuracy, and BWE.

Demand uncertainty (2008)	Forecast accuracy	Average BWE
lowest to highest	most accurate to least	least to greatest
Mid-size sedan	Mid-size sedan	Mid-size Pickup truck
Full-size pickup truck	Mid-size Pickup truck	Sports car
Wagon	Full-size sedan	Full-size SUV
Van	Full-size pickup truck	Full-size pickup truck
Full-size sedan	Van	Compact sedan
Mid-size pickup truck	Wagon	Mid-size sedan
Full-size SUV	Compact sedan	Wagon
Compact sedan	Full-size SUV	Mid-size SUV
Sports car	Sports car	Full-size sedan
Mid-size SUV	Mid-size SUV	Van

supply chain performance.

Table 6a
Pearson correlation test for all ten auto types.

SARIMA		No SARIMA	
Coefficient	*P*-value	Coefficient	*P*-value
−0.06073	0.5484	-0.08743	0.4125

Table 6b
Pearson correlation test for each auto type.

Type	SARIMA		No SARIMA	
	Coefficient	*P*-value	Coefficient	*P*-value
Compact sedan	−0.4940	0.1467	−0.5859	0.0974
Mid-size sedan	0.0463	0.8990	−0.3616	0.3390
Full-size sedan	−0.0265	0.9421	−0.0968	0.8043
Sports car	−0.8217	0.0035	−0.9812	< 0.0001
Mid-size SUV	0.0750	0.8368	0.0306	0.9377
Full-size SUV	−0.5811	0.0781	−0.6101	0.0810
Mid-size pickup truck	0.5295	0.1155	0.5055	0.1651
Full-size pickup truck	−0.7147	0.0202	−0.7514	0.0196
Van	0.6472	0.0431	0.3102	0.4165
Wagon	0.0684	0.8510	−0.0189	0.9614
Car	0.3792	0.2798	0.2316	0.5488
Truck	−0.3271	0.3563	−0.3786	0.3151
Light vehicle	−0.2221	0.5374	−0.2302	0.5513

4.4. The effect of aggregate forecasting

One of the major objectives of this study is to investigate the impacts of aggregate forecast on the BWE. Aggregation is believed to reduce the BWE (Cachon et al., 2007; Chen and Lee, 2012) and improve forecast accuracy. Our test results in Tables 2a, 2b, 4a and 4b seem to generally and plausibly support the previous wisdom

Table 7
Test results for aggregate forecast.

Car	BWE difference	Fisher's LSD	Tukey's HSD	Truck	BWE difference	Fisher's LSD	Tukey's HSD
Compact sedan - Car	0.2595			Midsize SUV - Truck	0.1176		
Midsize sedan - Car	0.5026	**		Fullsize SUV - Truck	−0.3094		
Fullsize sedan - Car	1.1328	***	***	Midsize pickup - Truck	−0.3380		
Sports car - Car	0.1235			Fullsize pickup - Truck	−0.2605		
Car - Light vehicle	0.1012			Wagon - Truck	0.0377		
				Van - Truck	0.6720	**	
				Truck - Light vehicle	0.5670	*	

*** is significant at 0.01 level.
** is significant at 0.05 level.
* is significant at 0.1 level.

and generate our interests to examine whether it is the aggregate forecasting that leads to the above positive findings. In order to verify the impacts of aggregate forecasting, we conduct two different tests to uncover the buried connections. We first investigate whether there exists an association between the forecast accuracy and the BWE at aggregate level and, then, test whether there exists significant improvement of the BWE from disaggregate level to aggregate level.

Our findings in Table 6b, unfortunately, does not support an accuracy-BWE association at the aggregate level since we observe insignificant Pearson correlation coefficients for Car, Truck, and Light vehicle. It implies the improved forecastability due to aggregate forecasting has zero influences on the BWE. Then, comparing the average BWE measure, we use both the Fisher's least significant different (LSD) and Tukey's honest significant difference (HSD) tests to evaluate the differences of the BWE measures in all three levels and report the results in Table 7.

For the cars, the BWE measure from the second level, Car, is significantly smaller than that of mid-size and full-size sedan. On the other hand, the difference of BWE between group and brand level is insignificant. For the trucks, the test results show BWE improvement from between Van and Truck and marginally between Truck and Light vehicle. These post hoc results indicate that the aggregate forecasting is not necessarily a solution to manage the BWE. In conclusion, we are unexpected to conclude that though aggregate forecasting lead to better forecast accuracy, it seems not to solve the BWE.

4.5. The effect of other forecasting-related factors

Given the conclusions of the correlation analysis, we further tested the impact of lead time and forecasting responsiveness on the BWE since forecasting accuracy does not necessarily lead to the BWE. Although our evidence implies that there is no or a negative relationship between the BWE and the forecast accuracy, we suggest that the safety stock created by forecast errors may be a major explanation for the creation of the BWE because the BWE measures obtained from perfect forecast always equals to one. If lead time is ignored, the perfect forecast is developed so that the variation of the order quantity is the same as the variation of the demand, then zero forecast error leads to zero safety stock and, hence, zero BWE (i.e., BWE$=$1). Given that this study aims to investigate the effects of forecast-related issues on the BWE, it is instructive to take up the safety stock problem.

There are two factors related to the safety stock that must be tested, namely, lead time and forecasting responsiveness. The test results are presented in Table 8. The duration of the lead time has impacts on the safety stock when the OUT policy is used (Krajewski et al., 2012). Meanwhile, the more responsive the forecast, the less safety stock the firm might need since it can quickly respond to the market change. Due to the issues tested and the complexity and variety of the forecasting models applied in this study, simple exponential smoothing (SES) forecasting was selected to conduct this sensitivity analysis. We consider three different lead time settings, three different numbers of periods considered in moving average forecast, and three different values of the SES parameter to investigate the impacts of lead time and forecasting responsiveness on the BWE. The three lead time settings are no lead time ($l=1$), one-period lead time ($l=2$), and two-period lead time ($l=3$). The three moving average periods are 3-period, 6-period, and 12-period (from responsive forecast to stable forecast) and the three smoothing parameters are 0.25, 0.5,

Table 8
Selected results on impacts of lead time and forecast responsiveness on the BWE.

Auto type	BWE Performance						
	Lead time	MA(3)	MA(6)	MA(12)	SES(.25)	SES(.5)	SES(.75)
Compact sedan	No lead time	1.756	1.324	1.213	1.375	1.907	2.336
	One period	2.853	1.942	1.348	1.980	3.074	3.832
	Two Period	3.848	2.433	1.496	2.700	4.116	5.832
Mid-size sedan	No lead time	1.897	1.568	1.421	1.636	2.222	3.023
	One period	3.346	2.303	1.528	2.405	4.262	6.937
	Two Period	5.324	3.220	1.648	3.354	7.087	12.685
Sports car	No lead time	1.988	1.462	1.148	1.627	2.486	3.478
	One period	2.819	1.976	1.284	2.331	4.317	6.672
	Two Period	3.707	2.574	1.428	3.164	6.251	10.624
Mid-size SUV	No lead time	2.307	1.409	1.167	1.694	2.801	4.204
	One period	3.794	1.809	1.358	2.477	4.948	7.629
	Two Period	5.079	2.272	1.571	3.319	6.953	11.635
Van	No lead time	2.363	1.735	1.387	2.047	2.932	3.991
	One period	3.767	2.289	1.650	2.993	5.417	8.954
	Two Period	5.629	2.996	1.953	4.149	8.786	16.165

and 0.75 (from stable forecast to responsive forecast). The BWE results presented in Table 7 showed that stable forecasts always lead to less BWE than the responsive forecasts when either the moving average or the simple exponential smoothing method is used to make forecasts. This finding is similar to that in Chen et al. (2000b) that "if a retailer periodically updates the mean and variance of demand based on observed customer demand data, then the variance of the orders placed by the retailer will be greater than the variance of demand." Accordingly, responsive forecasting though responds quickly to the change, it is also greatly affected by the sudden large scale unexpected events. Clearly, the quick response leads to a greater range of the order quantity and a stronger bullwhip effect. Likewise, the impact of lead time is the same as previously concluded that the longer the lead time, the greater the BWE (Chen et al., 2000b). The findings imply that it is the demand, instead of the forecast accuracy, that amplifies or reduces the BWE because if the demand tends to be stable, the forecast would be stable and the BWE can be reduced or remain unchanged. When considering both factors simultaneously, the magnitude of the BWE doubles when lead time increase from zero to two periods for a stable forecast, while the magnitude of the BWE increases by more than three times for a responsive forecast. As a result, our findings confirm previous knowledge regarding the impacts of lead time and responsive forecast upon the BWE, regardless of the demand uncertainty of the product.

4.6. Managerial implications for forecast accuracy

This study examines one common belief that a more accurate forecast on demand helps control the BWE (Duc et al., 2008; Jaipuria and Mahapatra, 2014). Numerous different popularly tested forecasting methods were considered in this real-world-data study to measure the BWE and the relationship between the BWE and forecast accuracy at three different product levels. Moreover, additional practices, such as aggregate forecasting, were applied to better improve forecast accuracy. The empirical results showed that the BWE always exists and that no or negative relationships between the BWE and the forecast accuracy were concluded. Moreover, although observing improved forecast accuracy from type-level forecast to brand-level forecast, we did not obtain convincing relationship between the BWE improvement and forecast accuracy for aggregate forecasting. In other words, even though aggregate forecast may lead to better forecast accuracy, the BWE performance at the aggregate (group or brand) level does not significantly differ from the BWE performance at the type level. On the other hand, these observations support some previous findings. For example, the BWE always exists, but choosing an appropriate forecasting method can alleviate the BWE (Wright and Yuan, 2008). On the other hand, the test results of zero or negative relationships raise our concern regarding the previous wisdom that improving forecast accuracy can control the BWE or improve the BWE performance.

In addition to the test results from ANOVA and correlation, the examination of the responsive forecast and stable forecast implicitly support the finding that improved forecast accuracy does not necessarily have a power to positively influence the BWE. Responsive forecast implies that the forecasted values respond to the variation of the actual values sooner than the stable forecast. From the sensitivity analysis results, it is obvious that as the forecasted values are more stable, the BWE is less significant. The observation partially shed light on the notion that it is not the forecast accuracy which enhances or alleviates the BWE but the attributes of a forecasting method do. With all of the findings, we believe that business managers in retailing businesses should be flexible in choosing the forecasting method. Even though a champion forecasting method can be identified at one product level, the same method is not necessarily the best method at a different product level. Finally, although the BWE can be controlled, it is hard to be removed completely unless the perfect demand forecast is attainable in every time period. Meanwhile, since the relationship between the forecast accuracy and the BWE is unsettled while lead-time is, business managers in retailing business are recommended to develop the capability of quick order fulfillment while using an appropriate stable forecast.

5. Conclusions

BWE is believed to reduce the supply chain profitability, increase the total supply chain inventory, and hurt the customer service level. Numerous efforts have been made to recognize the causes and solutions for the bullwhip phenomenon by using analytical-based research methods. This study, responding to Fisher (2007)'s suggestion, intends to empirically validate the analytical findings in previous bullwhip-related works. Developing on the foundations of several influential studies, such as Chen et al. (2000b), Zhang (2004), etc., this research validates the existing findings in the following ways.

First, forecasted demand is used to decide the order quantity and the BWE measures provide us a picture of how many times the variation of orders is greater than the variation of the demand. The BWE measures vary greatly among different forecasting techniques for the same auto type. Since the replenishment policy is fixed, the BWE measures show the effect of forecasting methods used and may suggest a good method for order management. Moreover, BWE measures from a few specific forecasting methods are consistently greater than those from others. Likewise, the ANOVA results further confirm the variation of BWEs with the choice of the forecasting method. These observations mean that if a less-desired forecasting technique is used, the BWE problem can be multiple times more serious than that from an alternative forecasting technique. Namely, if a firm can apply right forecasting technique for demand forecasting, it helps to control the problem of the BWE.

Next, the results regarding the forecast accuracy tests indicate that if the perfect forecast is unavailable, the BWE always exists when the OUT policy is used. The results show that as long as the forecast error exists, the safety stock is needed and the BWE is unavoidable. However, correlation tests showed zero association between the forecast accuracy and the BWE. Considering all obtained outcomes, the authors suggest that when the OUT policy is used, firms put efforts on managing the variation of the safety stock since the variation may be the major cause of the BWE, or that firms use other inventory policies. Moreover, lead time and the responsive forecast also enhance the scales of the BWE. The results imply that the firm should spend resources on minimizing the replenishment lead time and possibly stabilizing the demand variation. Finally, although aggregation may help to reduce the BWE as proposed by previous efforts, our findings show that aggregate planning is not necessarily helpful to control BWE. The findings further lead to the concern that aggregation is simply a fallacy which masks up the performance.

This study follows strictly the Box-Jenkin's approach to identify a feasible SARIMA model for each of the auto types considered and, then, uses simulation to empirically investigate the questions at issue. In addition, formal tests were carried out to verify the validity and reliability of the test results. However, a few limitations are unavoidable when carrying out the work. The first limitation comes from the data considered. Auto makers and dealers sell various versions for one specific auto model. For example, there are multiple versions in the Cobalt family as the type of the compact sedan. In this study, it is assumed that the product is homogenous in the market. This assumption is used in all previous analytical bullwhip studies but is far from what we observe in the market. The other limitation is the restriction for external generalization. Although the study is made based on the auto types in ten different light vehicle categories, it is hard to assume that the same results are expected if studying on other auto makers, such as BMW or Porsche. This limitation is hard to avoid since different auto makers may target at different customer groups and have different business strategies. Another limitation should be pointed out is that the BWE is rooted from many different causes. Although this study investigates the impact of the demand forecasting, there are additional factors which simultaneously affect the replenishment decision causing the BWE varied. For example, inventory policy may influence the order quantity and change the range/variation of the order quantity in the considered time period so that the BWE value changes accordingly. Although the three limitations are influential, they do not reduce the contributions of the current study because this study provides a systematic thinking and practice to investigate and evaluate the BWE-related issues. For future research, one possible extension is to empirically study the BWE on the supply chain level since this study emphasizes on the retailer echelon. A supply chain level study may enrich the current findings with a broader picture. Furthermore, additional demand-side factors may be taken into consideration to investigate the roles they play in the BWE and develop an interdisciplinary understanding to control the BWE. Finally, we believe that there is a need to investigate the dynamism among demand, inventory, and production in a supply chain to further investigate BWE problems.

Acknowledgments

The authors especially thank the anonymous referees for constructive comments that improved the rigor of the paper. The first author would like to thank the College of Business at Georgia Southern University for the generous research award that supported this research.

Appendix A. Other time-series forecasting methods

In addition to the SARIMA model, many other popularly applied time-series forecasting methods have been considered to investigate the effect of accurate forecasting in the supply chain inefficiency. The first forecasting model is the naïve model as shown in Eq. (A.1) below.

$$\hat{D}_t = D_{t-1} \tag{A.1}$$

Eq. (2) is a general form for a simple moving average forecasting model.

$$\hat{D}_t = \frac{\sum_{i=1}^{n} D_{t-i}}{n} \tag{A.2}$$

Eq. (3) presents a general form of a single exponential weighted moving average forecasting model.

$$\hat{D}_t = \alpha D_{t-1} + (1-\alpha)\hat{D}_{t-1} \tag{A.3}$$

Eq. (4) is a forecasting model using Brown's double exponential smoothing (DES) method (Hanke and Reitsch, 1991; Lawrence et al., 2008). Brown's DES method is one of the two major double exponential smoothing methods used popularly. This approach adds a smoothed value to measure the trend and is estimated as in Eq. (A.4).

$$\hat{D}_t = a_{t-1} + b_{t-1} \tag{A.4}$$

where

$$A_t = \alpha_b D_t + (1-\alpha_b)(A_{t-1})$$

$$A'_t = \alpha_b A_t + (1-\alpha_b)A'_{t-1}$$

$$a_t = 2A_t - A'_t$$

$$b_t = \frac{\alpha_b}{1-\alpha_b}(A_t - A'_t)$$

A_t: single exponential smoothed value of Z_t at time t,
A'_t: double exponential smoothed value of Z_t at time t,
α_b: smoothing constant ($0 < \alpha_b < 1$),
a_t: the smoothed value at the end of period t, and
b_t: the estimate of the trend at the end of period t.

The next forecasting technique is Holt's DES method which is shown as Eq. (A.5). This approach is different from Brown's DES in considering a trend by using two parameters, α and β (both are assumed to be between 0 and 1). The former parameter is used to smooth the level and the latter is for the trend. Holt's DES model presented in Eq. (A.5) below is composed of (1) the estimate of level, a trend-adjusted single exponential smoothed value, and (2) the estimate of trend.

$$\hat{D}_t = L_{t-1} + b_{t-1} \tag{A.5}$$

where

$$L_t = \alpha D_t + (1-\alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$$

L_t: the estimate of level,
b_t: the estimate of trend.

Seasonal exponential smoothing, present in Eq. (A.6), considers the influences of the seasonality in an exponential smoothing form. The m period prediction equation is as below.

$$\hat{D}_{t+m} = L_t + S_{t-p+m} \tag{A.6}$$

where

$$L_t = \alpha(D_t - S_{t-p}) + (1-\alpha)L_{t-1}$$

$$S_t = \delta(D_t - L_t) + (1-\delta)S_{t-p}$$

Damped exponential smoothing, shown in Eq. (A.7), considers the impact of the damped trend. The m period prediction equation is as below.

$$\hat{D}_{t+m} = L_t + \sum_{i=1}^{k} \varnothing^i T_t \tag{A.7}$$

where

$$L_t = \alpha D_t + (1-\alpha)(L_{t-1} + \varnothing T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)\varnothing T_{t-1}$$

The last forecasting method is Holt-Winters exponential smoothing with a seasonal component and a trend in multiplicative form. The m period ahead forecast is presented in Eq. (A.8) below.

$$\hat{D}_{t+m} = (L_t + m b_t)S_{t+m-s} \tag{A.8}$$

where

$$L_t=\alpha D_t/S_{t-s}+(1-\alpha)\left(L_{t-1}+b_{t-1}\right)$$

$$b_t=\beta\left(L_t-L_{t-1}\right)+(1-\beta)b_{t-1}$$

$$S_t=\gamma D_t/L_t+(1-\gamma)S_{t-s}$$

S_t: the multiplicative seasonal factor.

Appendix B. The Orders of the SARIMA models as specified by p, d, q, P, D, and Q

The procedure to estimate the orders of SARIMA model is explained here. In the beginning, we conducted unit root tests to verify whether the considered time series are stationary. Due to the unit root test results and observed seasonality, we decided to have a seasonal difference for 12 periods. Since the unit root test results approve the stationarity of the consider time series, the order for "d" was determined to be zero while the order for "D" was 12. Next, we used autocorrelation plot, partial-autocorrelation plot, and inverse-autocorrelation plot to determine the order for "p", "q", "P", and "Q". Using these plots is the standard approach described in Box et al. (1994) to identify the order for AR and MA terms. After a model was fit based on the interpretation of the plot, we further checked (i) the significance level of the coefficient, (ii) the residual plot, and (iii) the white noise plot for the accuracy and the eligibility of the model. As a matter of fact, since a few candidate models may be concluded through the mentioned process, we used AIC and SBC as the last selection criteria to determine the final model and the orders of the "p", "q", "P", and "Q". In addition, since we used SAS for model generation and forecast, we also relied on the online user's guide to develop the code/program and confirm the test results by a demonstration of a similar example available in SAS website. In short, each suggested univariate multiplicative seasonal time series model was estimated and tested by several standard procedures and tests requested in Box et al. (1994). The information for the example presented in SAS website can be visited from this web link (http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_arima_sect056.htm).

Then, from the estimated results of autocorrelation function and partial autocorrelation function of the sales, we identify the most fitted time series model for each selected Chevrolet auto type using the maximum likelihood method based on our own best effort. The proposed SARIMA models for the demand of each auto type are presented below in Eq. (B.1) (compact sedan), Eq. (B.2) (Mid-size sedan), Eq. (B.3) (Full-size sedan), Eq. (B.4) (sports car), Eq. (B.5) (Mid-size SUV), Eq. (B.6) (Full-size SUV), Eq. (B.7) (Mid-size pickup truck), Eq. (B.8) (Full-size pickup truck), Eq. (B.9) (Van), Eq. (B.10) (Wagon), Eq. (B.11) (Car), Eq. (B.12) (Truck), and Eq. (B.13) (Light vehicle).

$$\left(1-0.807B+0.120B^{11}\right)\left(1+0.647B^{12}+0.334B^{24}\right)\nabla_{12}^{1}D_t=\left(1-0.347B+0.219B^{14}\right)\epsilon_t \quad (B.1)$$

$$\left(1-0.585B+0.154B^{3}\right)\nabla_{12}^{1}D_t=\left(1-0.159B^{7}-0.770B^{12}\right)\epsilon_t \quad (B.2)$$

$$\left(1-0.630B+0.174B^{4}\right)\nabla_{12}^{1}D_t=\left(1-0.193B^{11}\right)\left(1-0.885B^{12}\right)\epsilon_t \quad (B.3)$$

$$\left(1-0.498B-0.217B^{4}\right)\left(1+0.677B^{12}+0.277B^{24}\right)\nabla_{12}^{1}D_t=\left(1+0.166B^{4}-0.228B^{10}-0.250B^{16}\right)\epsilon_t \quad (B.4)$$

$$(1-0.977B)\nabla_{12}^{1}D_t=\left(1-0.531B-0.200B^{2}+0.182B^{21}\right)\left(1-0.847B^{12}\right)\epsilon_t \quad (B.5)$$

$$\left(1-0.519B-0.291B^{3}-0.162B^{9}\right)\nabla_{12}^{1}D_t=\left(1-0.365B^{3}\right)\left(1-0.794B^{12}\right)\epsilon_t \quad (B.6)$$

$$(1-0.719B)\left(1+0.389B^{12}\right)\nabla_{12}^{1}D_t=\left(1-143B^{2}+0.226B^{14}\right)\epsilon_t \quad (B.7)$$

$$(1-0.339B)\left(1+0.461B^{12}+0.318B^{24}\right)\nabla_{12}^{1}D_t=\left(1+0.149B^{6}\right)\epsilon_t \quad (B.8)$$

$$\left(1-0.385B-0.195B^{3}-0.156B^{5}\right)\nabla_{12}^{1}D_t=\left(1+0.075B^{14}\right)\left(1-0.613B^{12}\right)\epsilon_t \quad (B.9)$$

$$\left(1-0.539B-0.349B^{4}\right)\nabla_{12}^{1}D_t=\left(1-0.812B^{12}\right)\epsilon_t \quad (B.10)$$

$$\left(1-0.392B-0.129B^{4}-0.129B^{6}\right)\nabla_{12}^{1}D_t=\left(1-0.845B^{12}\right)\epsilon_t \quad (B.11)$$

$$\left(1-0.317B-0.176B^{6}-0.234B^{9}-0.162B^{21}\right)\nabla_{12}^{1}D_t=\left(1-0.676B^{12}\right)\epsilon_t \quad (B.12)$$

$$\left(1-0.340B-0.172B^{6}-0.203B^{9}-0.162B^{21}\right)\nabla_{12}^{1}D_t=\left(1-0.606B^{12}-0.241B^{24}\right)\epsilon_t \quad (B.13)$$

The estimated coefficients are mostly significant at the 1% level except that the AR(11) term (−0.120) of the Compact sedan, MA (4) term (−0.166) of the Sports car, the MA(6) term (−0.149) of the Full-size pickup truck, AR(5) term (0.156) of the Van, and AR (4) and AR(6) terms of the "Car" are significant at the 5% level. Meanwhile, in order to attain the white noise requirement, the MA (2) terms of the Mid-size pickup truck has to be considered though it is significant at merely 90%.

References

Ali, M.M., Boylan, J.E., Syntetos, A.A., 2012. Forecast errors and inventory performance under forecast information sharing. Int. J. Forecast. 2012, 830–841.

Alwan, L.C., Liu, J.J., Yao, D.-Q., 2003. Stochastic characterization of upstream demand processes in a supply chain. IIE Trans. 35, 207–219.

Barlas, Y., Gunduz, B., 2011. Demand forecasting and sharing strategies to reduce fluctuations and the bullwhip effect in supply chains. J. Oper. Res. Soc. 62, 458–473.

Bayraktar, E., Koh, S.C.L., Gunasekaran, A., Sari, K., Tatoglu, E., 2008. The role of forecasting on bullwhip effect for E-SCM applications. Int. J. Prod. Econ. 113, 193–204.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis: Forecasting and Control, 3rd ed. Prentice Hall, Englewood Cliffs, NJ.

Cachon, G.P., Olivares, M., 2010. Drivers of finished-goods inventory in the U.S. Automobile Industry. Manag. Sci. 56, 202–216.

Cachon, G.P., Randall, T., Schmidt, G.M., 2007. In search of the bullwhip effect. Manuf. Serv. Oper. Manag. 9, 457–479.

Chandra, C., Grabis, J., 2005. Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand. Eur. J. Oper. Res. 166, 337–350.

Chen, F., Drezner, Z., Ryan, J.K., 2000a. The impact of exponential smoothing forecasts on the bullwhip effect. Nav. Res. Logist. 47, 269–286.

Chen, F., Drezner, Z., Ryan, J.K., Simchi-Levi, D., 2000b. Quantifying the bullwhip effect in a simple supply chain: the impact of forecasting. Manag. Sci. 46, 436–443.

Chen, L., Lee, H.L., 2012. Bullwhip effect measurement and its implications. Oper. Res. 60, 771–784.

Dejonckheere, J., Disney, S.M., Lambrecht, M., Towill, D.R., 2003. Measuring and avoiding the bullwhip effect: a control theoretic approach. Eur. J. Oper. Res. 147, 567–590.

Duc, T.T.H., Luong, H.T., Kim, Y.-D., 2008. A measure of bullwhip effect in supply chains with a mixed autoregressive-moving average demand process. Eur. J. Oper. Res. 187, 243–256.

Fildes, R., Kingsman, B.G., 2010. Incorporating demand uncertainty and forecast error in supply chain planning models. J. Oper. Res. Soc. 62, 483–500.
Fisher, M., 2007. Strengthening the empirical base of operations management. Manuf. Serv. Oper. Manag. 9, 368–382.
Forrester, J., 1958. Industrial dynamics: a major breakthrough for decision makers. Harv. Bus. Rev. 36, 37–66.
Gaalman, G., Disney, S.M., 2009. On bullwhip in a family of order-up-to policies with ARMA(2,2) demand and arbitrary lead-times. Int. J. Prod. Econ. 121, 454–463.
Geary, S., Disney, S.M., Towill, D.R., 2006. On bullwhip in supply chains - historical review, present practice and expected future impact. Int. J. Prod. Econ. 101, 2–18.
Gilbert, K., 2005. An ARIMA supply chain model. Manage. Sci. 51 (2), 305–310.
Hanke, J.E., Reitsch, A.G., 1991. Moving Averages and Smoothing Methods. Prentice-Hall, Englewood Cliffs, NJ.
Hosoda, T., Disney, S.M., 2009. Impact of market demand mis-specification on a two-level supply chain. Int. J. Prod. Econ. 121, 739–751.
Jaipuria, S., Mahapatra, S., 2014. An improved demand forecasting method to reduce bullwhip effect in supply chains. Expert Syst. Appl. 41, 2395–2408.
Klug, F., 2013. The internal bullwhip effect in car manufacturing. Int. J. Prod. Res. 51, 303–322.
Krajewski, L.J., Ritzman, L.P., Malhotra, M.K., 2012. Operations Management: Processes and Supply Chain, 10 ed. Pearson, United Kingdom.
Lawrence, K., Klimberg, R.K., Lawrence, S.M., 2008. Fundamentals of Forecasting Using Excel, 1st ed. Industrial Press, Inc., United States.
Lee, H.L., Padmanabhan, V., Whang, S., 1997a. The bullwhip effect in supply chains. Sloan Manag. Rev. 38, 93–102.
Lee, H.L., Padmanabhan, V., Whang, S., 1997b. Information distortion in a supply chain: the bullwhip effect. Manag. Sci. 43, 546–558.
Li, Q., Disney, S.M., Gaalman, G., 2014. Avoiding the bullwhip effect using Damped Trend forecasting and the Order-Up-To replenishment policy. Int. J. Prod. Econ. 149, 3–16.
Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications. Int. J. Forecast. 16, 451–476.
McCarthy, T.M., Davis, D.F., Golicic, S.L., Mentzer, J.T., 2006. The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices. J. Forecast. 25, 303–324.
Sterman, J.D., 1989. Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. Manag. Sci. 35, 321–339.
Syntetos, A.A., Nikolopoulos, K., Boylan, J.E., 2010. Judging the judges through accuracy-implication metrics: the case of inventory forecasting. Int. J. Forecast. 26, 134–143.
Waller, B., 2004. Market responsive manufacturing for the automotive supply chain. J. Manuf. Technol. Manag. 15, 10–19.
Wright, D., Yuan, X., 2008. Mitigating the bullwhip effect by ordering policies and forecasting methods. Int. J. Prod. Econ. 113, 587–597.
Zhang, X., 2004. The impact of forecasting methods on the bullwhip effect. Int. J. Prod. Econ. 88, 15–27.
Zotteri, G., 2013. An empirical investigation on causes and effects of the Bullwhip-effect: evidence from the personal care sector. Int. J. Prod. Econ. 143, 489–498.