Lancaster University Management School Working Paper 2006/016

The impact of forecasting on the bullwhip effect

Ioannis Stamatopoulos, Ruud Teunter and Robert Fildes

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/publications/
LUMS home page: http://www.lums.lancs.ac.uk/

The impact of forecasting on the bullwhip effect

Ioannis Stamatopoulos, Ruud Teunter and Robert Fildes

Lancaster University Management School
LA1 4YX Lancaster, UK
fax +44 1524 844885
ioannis.stamatopoulos@dwp.gsi.gov.uk, r.fildes@lancaster.ac.uk, r.teunter@lancaster.ac.uk

March 8, 2006

Abstract

There has been strong empirical evidence that demand variability increases as one moves up the supply chain (from the retailer to the raw materials supplier), a phenomenon called bullwhip effect. This paper examines the bullwhip effect and in particular one of its main causes, demand forecasting. Key observations for the studies that deal with the impact of forecasting on the bullwhip effect are that: (1) all allow negative demands as well as negative orders for analytical tractability and (2) none considers the best exponential smoothing forecast without a prefixed smoothing constant. This paper validates the main findings in the literature when negative demands and negative orders are not allowed, using simulation. The main contribution is the inclusion of 'best' exponential smoothing as a forecasting method. This method is shown to explain some structural differences in bullwhip effect that have been observed in comparisons between naïve exponential smoothing and optimal forecasting. Therefore, it provides an important alternative to naïve smoothing for use in practice, especially as it is included in some of the more modern Demand Planning Systems.

Keywords: bullwhip; supply chain; forecast mis-specification

1. Introduction and Related Literature

The existence of the bullwhip effect has been acknowledged in a plethora of studies. In the context of industrial Dynamics, Forrester (1961) was the first to illustrate that it is common for the variance of perceived demand to the manufacturer to far exceed the variance of consumer demand and for seasonality to be larger for manufacturers than for retailers.

Caplin (1985) considered a retailer who follows a continuous review (s, S) inventory policy and proved that if the demands faced by the retailer are i.i.d., then the variance of the orders placed by the retailer is greater than the variance of the customer demand observed by the retailer and that the variance of the orders increases linearly in the size of the orders. Blinder (1986) documented the bullwhip effect in 20 different sectors of the economy, explaining it from a macro-economic perspective.

Kahn (1987) demonstrated the existence of the bullwhip effect when the retailer follows an optimal inventory policy and either demand in each period is positively serially correlated or the backlogging of excess demand is permitted. Sterman (1989) reported evidence of the bullwhip effect in the “Beer Distribution Game”. The experiment involved a supply chain with four players who made independent inventory decisions without consultation with other chain members, relying only on orders from the neighboring player as the sole source of communications. Under the linear cost structure, the experiment showed that the variances of orders amplified as one moved up the supply chain adding further confirming evidence for the bullwhip effect.

The bullwhip effect has been assigned to various causes, most of which come straight out of the Systems Thinking / Systems Dynamics discipline. According to Forrester (1961) the principal cause of the effect lies in the difficulties involving the information feedback loop between companies. In a similar manner, Sterman (1989) interpreted the phenomenon as a poor decision making based on a lack of understanding for the system as a whole. Blackburn (1991) indicated that the overriding cause of the bullwhip effect is the time delays between the supply chain echelons. Empirical observation of managerial behaviour in the beer distribution game led Senge and Sterman (1992) to blame it on lack of ‘system thinking’ by management.

Lee et al. (1997a, 1997b) followed a different approach and moved one step further by modeling the causes analytically. In their eyes, the existence of the phenomenon lies in the institutional and inter-organizational infrastructure and processes rather than the lack of full rationality and misperceptions of the members. They identified four sources of the bullwhip effect:

(i) demand forecast updating;

(ii) rationing game;

(iii) order batching and;

(iv) price fluctuations.

This paper focuses solely on the effect of demand forecast updating, which can succinctly be described as overreactions to changes in observed demand when making demand forecasts. We will therefore restrict ourselves to this source of the bullwhip effect in the remainder of the section.

Using a first order autoregressive (AR1) demand generation process, Chen et al. (2000a) studied the role of moving average forecasts in relation to the bullwhip effect in a two stage serial supply chain, where both members use an order-up-to policy. The authors illustrated that the increase in variability from the retailer to the manufacturer is a function of three parameters: the number of observations used in the moving average (p), the lead-time, and the demand correlation (ρ). More specifically, when p is large the increase in variability was trifling. In contrast, when p is small, there is a significant increase in the variability. Hence the smoother the demand forecasts, the smaller the increase in variability. In addition, order variability was found to be an increasing function of the lead-time parameter and a decreasing function of the correlation parameter.

In a sequel, Chen et al. (2000b) extended their results to the case where simple exponential smoothing is applied to forecast lead time demand generated from an AR1 process. Using analytical modeling they derived a lower bound on the variance of the orders placed by the retailer relative to the variance of customer demand. This bound was shown to produce similar results to those reported in the original study with relative variable sensitive to the same parameters. However their forecasting model (exponential smoothing) is mis-specified and not optimal for the underlying demand generation process.

Graves (1999) notes that exponential smoothing is not the optimal forecasting method for AR1 demand (in terms of minimizing the mean square errors). So, he instead analyzes the integrated MA process (0,1,1), for which exponential smoothing is optimal. He quantified the bullwhip effect for this type of demand generation process under an adaptive base stock policy.

Alwan et al. (2003) and Zhang (2004) also realize that exponential smoothing is not the optimal forecasting method for AR1 demand. However, instead of assuming a different demand process, they consider the optimal forecasting method for AR1 demand. The key differences they find comparing moving average and exponential smoothing, concern the effect of the demand correlation. The bullwhip no longer strictly increases in ρ, but first increases and then decreases. Furthermore, for negatively correlated demand, optimal forecasting completely eliminates the bullwhip effect and even turns it into a 'de-whip' effect.

All above discussed results on the impact of demand forecasting on the bullwhip effect are based on the assumption that demand as well as orders can be negative. These assumptions are unrealistic and made for analytical tractability only. The first contribution of this paper is that we drop these assumptions and validate the findings in the literature using simulation. In doing so, we also assume that demands not immediately satisfied are lost, whereas previous studies assumed backordering. It has often been recognized in the inventory literature that assuming lost sales is more realistic than assuming backordering, especially for retailers, but backordering models are usually preferred for analytical tractability.

The second and main contribution of this paper is that we also examine the effect of using the 'best' exponential smoothing method. All previous studies, when using exponential smoothing, employed a naïve method with a fixed smoothing constant. The best exponential smoothing method uses the constant that minimizes the mean square error. We remark that the method is often referred to as "MSE optimal exponential smoothing", but we use "best exponential smoothing" to avoid confusion with the (MSE) optimal forecasting method. The results will show that the differences in both the size of bullwhip effect as well as its behavior with respect to ρ, are indeed partially explained by the use of the naïve rather then the best exponential smoothing method.

The remainder of the paper is organized as follows. In Section 2, we describe the model and the simulation approach. In Sections 3-5, naïve exponential smoothing, best exponential smoothing and optimal forecasting are discussed, respectively. We end with a discussion, conclusions and directions for future research in Section 6.

2. Model and simulation details

We assume a simple two stage supply chain comprising of a single retailer at the lowest echelon, and a single manufacturer. We consider a single product. The demand per time unit faced by the retailer is stochastic and follows an AR1 process. That is,

$$D_t = \mu + \rho D_{t-1} + \varepsilon_t$$

where D_t and D_{t-1} denote demands in periods t and t-1, respectively, μ > 0, the autoregressive coefficient ρ is restricted to lie between -1 and 1, and the error terms ε_t are white noise (independent and identically distributed over time with mean 0 and variance σ^2).

If an order is placed in some period t, then it will arrive at the beginning of period t + L. Here, L can be interpreted as the lead time *including* the one period review time. Including the review time is without loss of generality and for simplicity of notation only, as order-up-to levels will now be based on demand forecasts for the next L rather than L + 1 time units.

In each period t, after observing demand, the retailer updates his forecast for lead time demand. The updating process depends on the forecasting method used. For exponential smoothing, the lead time demand forecast is $\hat{D}_t^L = L\hat{D}_t$, where $\hat{D}_t = \alpha D_t + (1-\alpha)\hat{D}_{t-1}$ is the updated 'flat' per period forecast. For optimal forecasting, the lead time demand forecast is $\hat{D}_t^L = \dfrac{L(1-\rho)-\rho(1-\rho^L)}{(1-\rho)^2}\mu + \dfrac{\rho(1-\rho^L)}{1-\rho}D_t$ (see Alwan et al. (2003)).

The retailer then calculates the order-up-to level for period t as follows:

$$S_t = \hat{D}_t^L + z\ \hat{\sigma}_{e,t}^L$$

where $\hat{\sigma}_{e,t}^L$ is an estimate of the standard deviation of the L period forecast error and z is the safety factor chosen to meet a desired service level. For all our simulation experiments the safety factor has been set to 1.64 corresponding to a 95% service level. Note that $\hat{\sigma}_{e,t}^L$ is not simply an estimate for the standard deviation of lead time demand, since the sampling error needs to be accounted for as well (Nahmias, 2001). For most inventory models this can be empirically estimated by $\hat{\sigma}_{e,t}^L = L^c\ \hat{\sigma}_{e,t}$ or simply $\hat{\sigma}_{e,t}^L = \sqrt{L}\ \hat{\sigma}_{e,t}$ under the assumption that the forecast errors are independent over time.

An important difference with previous research is that we assume negative orders are not allowed (see Section 1). So, an order is only placed if the inventory position is below the order-up-to level. Another difference is that we assume lost sales instead of backordering, i.e. demands that are not satisfied immediately are lost.

The bullwhip effect is measured by the ratio of the variance in retailer orders and the variance in retailer demands. The latter is easily shown to be $\sigma^2 / (1-\rho^2)$.

<u>Simulation details</u>

For each of the experiments presented in this chapter the simulation period is 10.000. The first 200 observations serve as the warm up period for our simulation and are excluded from the analysis. The number of runs is 100 and reported results are averages over all runs. Demand series are restricted to be positive by resampling negative observations. In all scenarios the error term variance σ^2 varies depending on the correlation coefficient so as to give a corresponding coefficient of variation of 0.2. According to the literature (Fildes and Beard, 1992) such demand patterns are commonly met in practice for a large number of fast moving items.

The simulation code is written in Matlab and the selection of the optimal smoothing constant for each simulation run is done using built in non linear optimization functions.

3 Exponential smoothing with a prefixed smoothing constant

In this section we test Chen's findings, later confirmed by other analytical studies, about the effects of the smoothing constant α, the autocorrelation coefficient ρ, and the lead time L. We start with α and ρ. Five different levels (0.1, 0.3, 0.5, 0.7, 0.9) are considered for α and three levels (0.5, 0, 0.5) for ρ, giving 15 experiments in total. The results are depicted graphically in Figure 1.

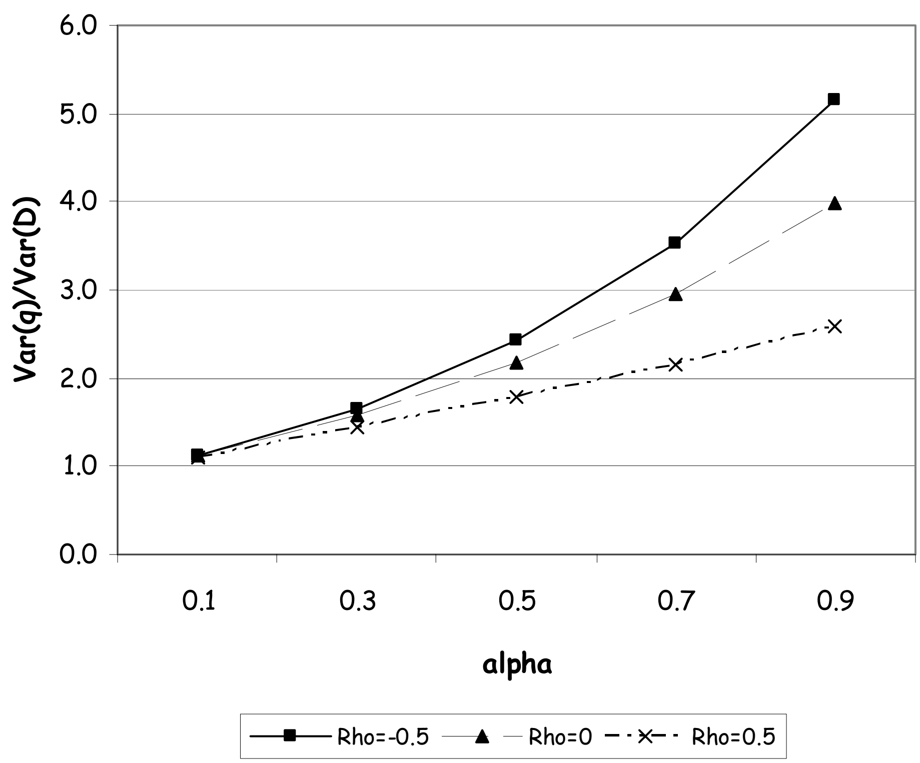

Figure 1 Bullwhip Effect for different values of α and ρ (L=1, CV=0.2)

These results validate Chen's findings that the bullwhip effect is increasing in α and decreasing in ρ. Next we explore (see Figure 2) the impact of lead times on the bullwhip effect for positively correlated, negatively correlated and white noise demand. For all autocorrelation demand patterns, the bullwhip effect is increasing in the lead time Again this is in line with Chen et al. Lead times magnify the increase in variability due to demand forecasting. The longer the lead time, the larger the inventory level required to deal with demand uncertainty. Hence if the retailer updates his target inventory level in each period, then longer lead times will trigger larger changes in the target inventory level resulting in higher volatility for the orders placed by the retailer.

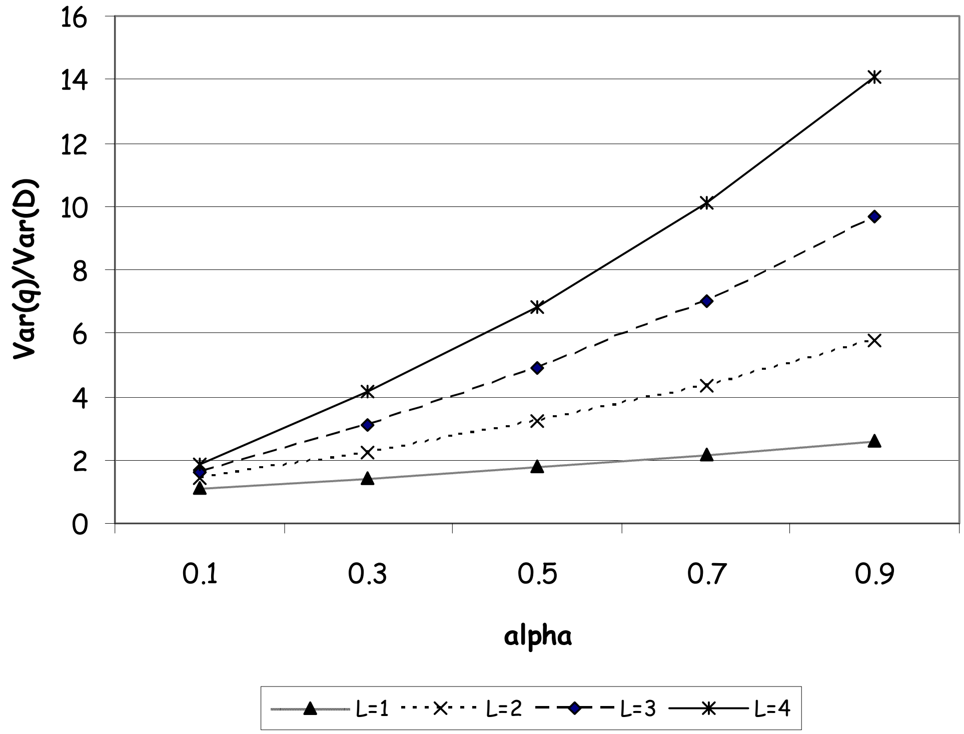

Figure 2 Bullwhip Effect for different values of L and α (ρ=0.5 CV=0.2 Z=1.64)

While these findings are certainly useful and shed light on the impact of forecasting on the bullwhip effect, they suffer from the limitation that the smoothing constants are set arbitrarily and not optimally. In practice, when the smoothing constants are set arbitrarily this is usually done following suggestions made by forecasting textbook authors. For instance, for simple exponential smoothing, Brown (1963) recommends a smoothing constant of 0.1. Smoothing constants between 0.1 and 0.3 are also frequently suggested in the practitioner forecasting literature.

Although some forecasting programs expect the value of the smoothing constant to be defined by the user, the more modern Demand Planning Systems have a core optimization routine in place, thus saving the practitioner from possible erroneous guesswork and operationalising aspects of best forecasting practice (Fildes et al., 1998). In the next section, we will therefore consider exponential smoothing with an optimized smoothing constant, which we refer to as best exponential smoothing (see Section 1).

4 Best exponential smoothing

When forecasting with exponential smoothing models, the optimization of the smoothing constant is often pivotal in forecast accuracy and should not be taken lightly. Empirical evidence (see Fildes et al., 1998) suggests that 'the performance of the smoothing methods depends on how the smoothing parameters are estimated'. It is also found that 'optimization (either at each time origin or at the first time origin) is shown to be superior to arbitrary (literature based) fixed values'. Computer simulation allows us to embark on the best (minimum MSE) exponential moving scheme rather than a simplistic arbitrary selection of the smoothing constant.

Figure 3 shows the effects of the autocorrelation coefficient ρ and the lead time L on the bullwhip effect when demand forecasts are estimated using best exponential smoothing.

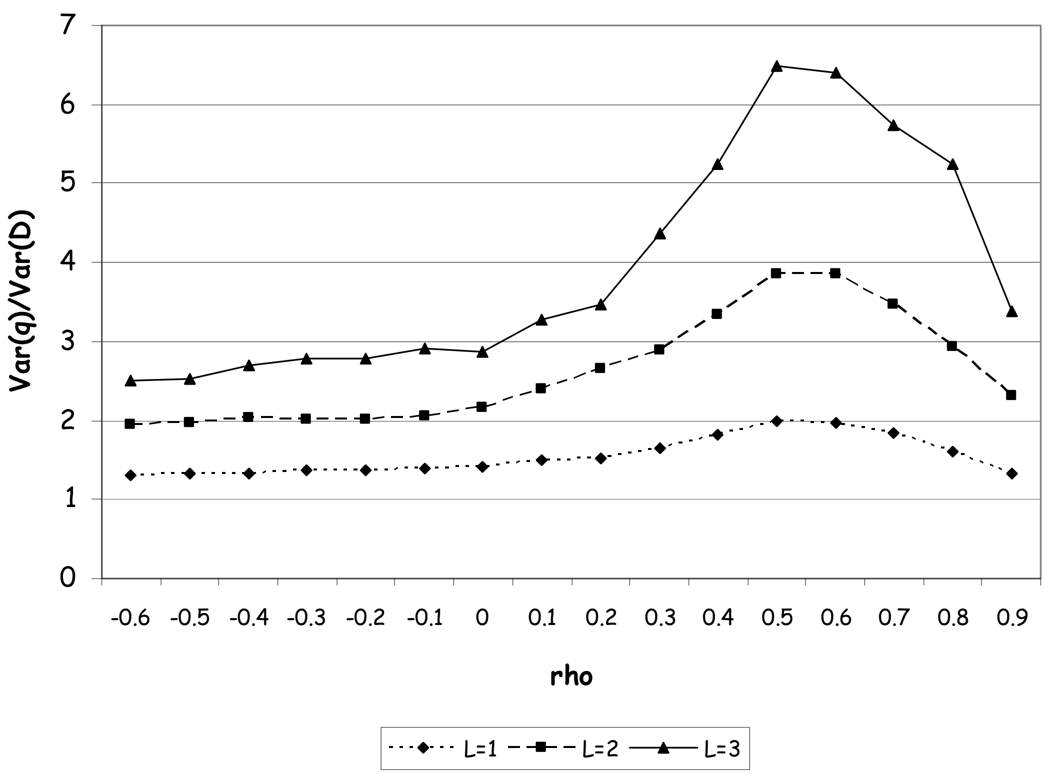

Figure 3 Bullwhip Effect for different values of ρ and L using an optimal exponential smoothing model (CV=0.2, Z=1.64)

An increase in the lead time still leads to an increased bullwhip effect over the entire range of demand correlation values considered. An obvious difference with the results for naïve exponential smoothing, however, is that the bullwhip effect is no longer strictly decreasing in ρ, but first increases from ρ = -1 up to about ρ = 0.5 before it starts to decrease. Recall from the introduction that the same change in behaviour has been observed in the literature when going from naïve exponential smoothing to optimal forecasting. What our results show is that going to from naïve to best optimal smoothing also achieves this change in behaviour. The change is certainly important, since it is driven by a large drop in bullwhip effect for smaller values of ρ. The way that best exponential smoothing achieves this drop is simply by using a smaller smoothing constant for smaller ρ. This is shown in Figure 4.

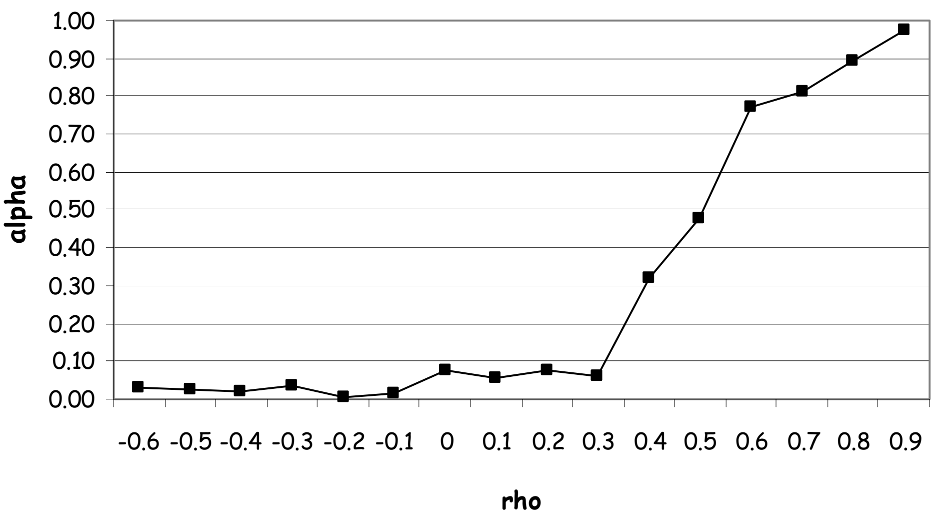

Figure 4 Optimal smoothing constant.

However, despite reducing the bullwhip effect considerably, it is not able to completely eliminate the bullwhip effect. As was analytically shown in the literature (see section 1) and will be validated in the next section, optimal forecasting *is* able to eliminate and can even turn it into a de-whip effect.

5 Optimal forecasting

To most researchers and practitioners, the forecast error is the difference between the actual and the forecast value. As Fildes and Kingsman (2005) have demonstrated few however have realized that this combines the randomness in the demand generation process (the error term) and the errors arising from not using the optimal forecasting model (forecast mis-specification).

As recognized by Alwan et al. (2003) and Zhang (2004) the only forecasting method that is truly optimal for the AR1 demand generation process is the ARIMA (1,0,0). Our simulation study offers us the unusual luxury for a forecaster, of knowing the underlying AR(1) parameters μ, ρ and σ_e^2, thus enabling us to apply the optimal forecasting model for the corresponding autoregressive process. Hence the only source of uncertainty in the demand forecasting process remains the random variation in demand itself. In practice however, such a high level of forecast precision and accuracy is not realistic as there would be inevitable errors in identifying and estimating the parameters of the demand generation process.

Figure 5 shows the bullwhip effect for different values of ρ and L using an optimal AR1 forecasting model.

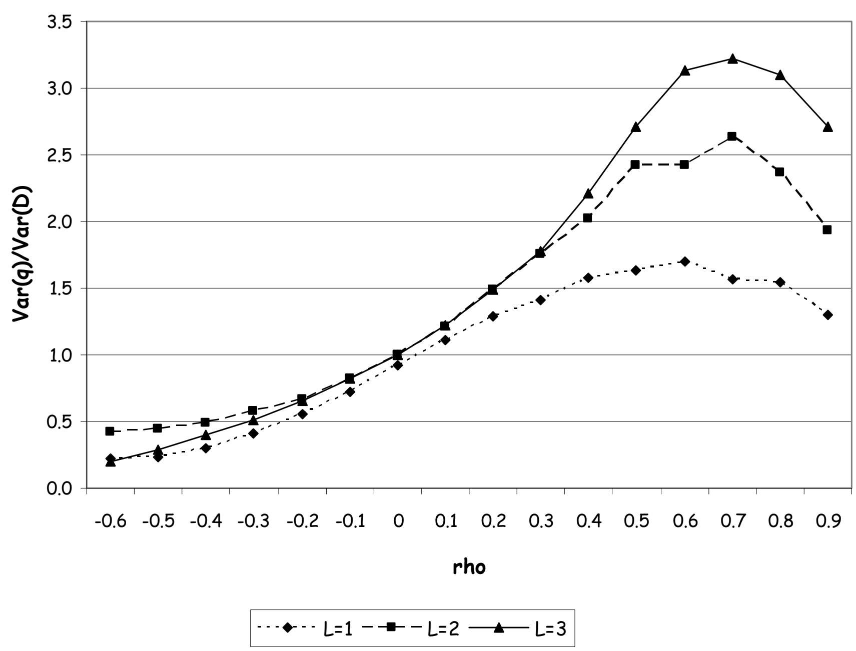

Figure 5 Bullwhip Effect for different values of ρ and L using an optimal AR1 forecasting model (CV=0.2 Z=1.64)

The pattern is similar to what we observed for the optimal exponential smoothing model. However, the bullwhip effect is further decreased over the entire range of ρ. In particular, optimal forecasting completely eliminates the bullwhip effect for negatively correlated demand and turns it into a de-whip effect. As a result, an increase in lead time may lead to a decrease in bullwhip for negatively correlated demand.

6. Discussion and conclusions

The main findings in the literature on the bullwhip effect for naïve (fixed constant) exponential smoothing and optimal AR1 forecasting have all been validated under the realistic assumption that neither negative demand nor negative orders are not allowed. In particular, where reduced demand correlation always leads to an increased bullwhip effect for naïve exponential smoothing, this does not hold for optimal forecasting. Indeed, optimal forecasting is able to eliminate the bullwhip effect for negatively correlated demand.

From the practitioner's point of view, that implies that optimal ARIMA forecasting has the potential of fully mitigating the bullwhip effect for companies with only a few big customers where a high demand in one period, when many of the customers have ordered, is likely to be followed by a lower demand in the following period. Price switches from high to low and vice versa as a result of regular promotions, can also be accountable for such oscillatory negatively correlated demand patterns.

Best exponential smoothing provides an interesting alternative to optimal forecasting. Though it is not able to eliminate the bullwhip effect, it does result in a considerable reduction compared to naïve exponential smoothing. Indeed, it shows the same structural pattern of a bullwhip that first increases (until a factor ρ of about 0.5) and then decreases with demand correlation.

An important practical advantage of best exponential smoothing over optimal forecasting is that it is included in modern Demand Planning Systems. Indeed, in realistic operational scenarios with hundreds of stock keeping units to be forecasted, the practitioner is likely to lack the expertise or the resources (whether time or specialized time series modelling software) to embark on the model identification, parameter estimation and tests of model adequacy that optimal forecasting requires. The easy alternative for him would be to apply an exponential smoothing model with its parameters automatically optimized by the forecasting software.

In addition, evidence from the M forecasting competitions (e.g. Makridakis and Hibon 2000), large scale empirical studies that examined forecasts for hundreds of time series, suggests that exponential smoothing is a difficult benchmark to beat. Indeed simple extrapolative methods such as our "best" exponential smoothing tend to offer comparable or better post sample forecast accuracy than the most sophisticated methods, including ARIMA models such as the one examined here. The reason is that time series parameters change over time, a characteristic which is especially true for supply chain data which often has little structure and relatively high degree of randomness. Therefore, having a sophisticated model that better fits historical data would not necessarily guarantee a more accurate post-sample forecast.

There are two important directions for further research. The first is to study more realistic multi-echelon supply chain settings. This study was limited to a two stage serial supply chain model, to ensure consistency with analytical models in the literature and allow a comparative study. However, research should now move forward to more complex chains. Another direction for further research is to include other sources of the bullwhip effect as well. For instance, the order batching effect could be included by considering order level, order-up-to level (s,S) policies.

References

Alwan LC, Liu JJ, Yao D. Stochastic Characterization of upstream demand processes in a Supply Chain. IEE Transactions 2003; 35; 207-219.

Blackburn J. Time-Based Competition - The Next Battleground in American Manufacturing. Business One: Irwin; IL; 1991.

Box GEP, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. Prentice Hall: 3rd ed.; Englewood Cliffs; N.J.; 1994.

Brown RG. Smoothing, Forecasting and Prediction. Prentice-Hall: Englewood Cliffs, NJ; 107–108; 1963.

Blinder A. Can the production smoothing model of inventory behavior be saved? Quarterly Journal of Economics 1986; 101; 431-53.

Caplin AS. The Variability of Aggregate Demand with (s, S) Inventory Policies. Econometrica 1985; 53; 1396-1409.

Chen F, Drezner Z, Ryan J, Simchi-Levi D. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. Management Science 2000a; 46 (3); 436 –443.

Chen F, Ryan J, Simchi-Levi D. The impact of exponential smoothing forecasts on the bullwhip effect. Naval Research Logistics 2000b; 47(4); 271 –286.

Fildes R, Beard C. Forecasting systems for production and inventory control. International Journal of Operations and Production Management 1992; 12; 4-27.

Fildes R, Hibon M, Makridakis S, Meade N. Generalising about univariate forecasting methods: further empirical evidence. International Journal of Forecasting 1998; 14; 339-358.

Fildes R, Kingsman BG. Demand Uncertainty and lot sizing in manufacturing systems: The effects of forecasting errors and mis-specification. Lancaster University Working Paper 2005.

Forrester JW. Industrial Dynamics. Cambridge: MIT Press; 1961.

Graves S. A single item inventory model for a nonstationary demand process. Manufacturing and Service Operations Management 1999; 1; 50-61.

Hamilton J. Time Series Analysis. Princeton: Princeton University Press; 1994.

Kann JA. Inventories and the volatility of production. The American Economic Review 1987; 667-679

Lee HL, Padmanabhan V, Whang S. The Bullwhip Effect in Supply Chains. Sloan Management Review 1997a; 38(3); 93-102.

Lee HL, Padmanabhan V, Whang S. Information Distortion in a Supply Chain: The Bullwhip Effect. Management Science 1997b; 43(4); 546-558.

Lee H, So K, Tang C. The value of information sharing in a two-level supply chain. Management Science 2000; 46(5); 626-643.

Makridakis S, Hibon M. The M3-Competition: results, conclusions and implications. International Journal of Forecasting 2000; 19; 1; 5-22.

Nahmias S. Production and Operations Analysis. Irwin/McGraw Hill: Fourth Edition; Burr Ridge Illinois; 2001.

Senge PM, Sterman JD. Systems Thinking and Organizational Learning. European Journal of Operational Research 1992; 59; 137-150.

Sterman J. Modelling managerial behaviour: Misperceptions of feedback in a dynamic decision making experiment. Management Science 1989; 35(3); 321-339.

Zhang X. The impact of forecasting methods on the bullwhip effect. International Journal of Production Economics 2004; 88; 15-27.