ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/2874/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Disney, Stephen Michael , Farasyn, I, Lambrecht, M., Towill, Denis Royston and Van de Velde, W. 2006. Taming the bullwhip effect whilst watching customer service in a single supply chain echelon. European Journal of Operational Research 173 (1) , pp. 151-172. 10.1016/j.ejor.2005.01.026

Publishers page:

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Disney, S.M., Farasyn, I., Lambrecht, M., Towill, D.R. and Van de Velde, W., (2006), "Taming the bullwhip effect whilst watching customer service in a single echelon of a supply chain", European Journal of Operational Research, Vol. 173, pp151–172. DOI: 10.1016/j.ejor.2005.01.026.

Taming the bullwhip effect whilst watching customer service in a single supply chain echelon

Disney, S.M.[1], Farasyn, I.[2], Lambrecht, M[3], Towill, D.R.[1] and Van de Velde, W.[2]

1. Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK. E-mail: disneysm@cardiff.ac.uk
2. Procter and Gamble Services Company NV, Temselaan 100, B-1853 Strombeek-Bever, Belgium. E-mail: vandevelde.wm@pg.com, farasyn.i@pg.com
3. Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B3000 Leuven, Belgium. E-mail: marc.lambrecht@econ.kuleuven.ac.be

---

We study a generalised order-up-to policy that has highly desirable properties in terms of order and inventory variance and customer service. We quantify exactly the variance amplification in replenishment orders, i.e. the bullwhip effect, and the variance of inventory levels over time, for i.i.d. and the weakly stationary Auto Regressive (AR), Moving Average (MA) and Auto Regressive Moving Average (ARMA) demand processes. We demonstrate that high customer service as measured by fill rate, and smooth replenishments need not increase inventory cost substantially. We observe that in some instances of the ARMA demand pattern this comes at the expense of a relatively small increase in safety stock, whilst in other instances inventory levels can actually be reduced.

(Order-up-to policies; ARMA demand; Bullwhip; Inventory variance; Fill rates)

---

1. Introduction

Inventory is used to provide a service to the customer, i.e. to give an immediate source of supply, and to buffer the production system from fluctuations in demand, but the prevalence of the bullwhip effect (Lee Padmanabhan and Whang, 1997a and b) suggests that this benefit is often lost (Baganha and Cohen, 1998). Both marketing and production executives clearly have a fundamental interest in inventory. Marketing will wish to set inventory levels to ensure a suitable customer service level. Additionally, production executives wish to use the inventory to reduce the "strain" on the production system resulting from the uncertain demand. These strains manifest themselves as lost capacity, transport and subcontracting premiums, over and under-time labour, training, and quality issues. The financial implications of these problems are often hard to determine, although academia and management have realised for a long time the benefit of designing replenishment rules to control inventory levels, maintain customer service and set production targets, see Magee (1956) and, Deziel and Eilon (1967) and Graves (1999). In reality a qualitative judgement is often taken by executives to manage the trade-off between the bullwhip effect, inventory holding and backlog costs whilst aligning the service proposition to the market place.

Inventory managers must consider two primary factors when making replenishments. First, a replenishment rule has an impact on order variability (as measured by the bullwhip effect, i.e., the ratio of the variance of orders over the variance of demand) shown to the supplier. Second, the replenishment rule has an impact on the variance of the net stock (as measured by the net stock amplification, i.e., the ratio of net stock variance over the variance of demand). The bullwhip effect mainly contributes to upstream costs, while the variance of net stock determines the stage's ability to meet a service level in a cost-effective manner. This is the key trade-off faced by a single-stage member of a supply chain. This is the key issue addressed in this paper. It is interesting to note that the problem described above may lead to non-cooperative behaviour.

Indeed, the bullwhip effect is driving costs at the upstream stage (e.g. the manufacturer or supplier) and consequently, the downstream stage (e.g. the retailer) may not worry about it. Even worse, dampening the bullwhip effect may have a negative impact on customer service. Why should a downstream stage be concerned with upstream costs? One answer becomes clear if the lead-time is also considered. By dampening the order variability, the upstream supply chain member will most likely be able to offer shorter lead times and this of course is beneficial for the downstream stage. However, the impact of dampening order variance on lead-time is not considered in this paper. Our referee has also suggested another reason why a downstream player should be concerned about upstream costs; the supplier may approach the downstream player with a co-ordination incentive. That is, the supplier may be able to reduce costs so much that they can offer to offset the downstream players increased costs.

The replenishment rule contributes to both the inventory levels and the production rates[1] of a product. The key to designing a good replenishment rule is to balance the inventory and production costs whilst ensuring a customer service level. In simple terms it is easy to visualise that we can "buy" a smoothed production level and customer service level with inventory. This is how inventory is supposed to be used, but often an improperly designed replenishment rule will destroy this beneficial inventory effect. General solutions are also hard to provide as inventory and production costs and service level requirements vary widely in practice. Furthermore, the trade-off also depends very much on the demand pattern.

There are two basic types of inventory replenishment rules; continuous time, fixed order and periodic re-ordering systems, Magee (1956) and Rao (2003). Fixed order systems result in the same quantity (or multiples thereof) of product being ordered at varying time intervals. In periodic systems a variable amount of product is ordered at regular, repeating intervals and the decision maker has to determine an order-up-to (OUT) level in each period. At fixed time intervals he compares the inventory position with the order-up-to level, and orders the difference. Given the common practice in retailing to replenish inventory frequently (e.g. daily), and the fact that OUT policies generate regular repeating schedules of inventory replenishments, we chose to model a periodic re-ordering system. As the decision operates at fixed time intervals, we may use the z-transform as a modelling and analysis tool (see Disney and Towill (2002)). In control systems engineering, the transfer function of a system represents the relationship describing the dynamics of the system under consideration. It algebraically relates a system's output (replenishment order) to its input (the demand process) and is defined as the ratio of the z-transform (the discrete time, z-domain, equivalent of the Laplace transform) of the output variable to the z-transform of the input variable. It is possible to develop a transfer function for every linear replenishment rule. We model time in integer multiples of a planning period, but our results will provide valuable insight regardless of the actual length of the planning period used in a particular application.

We have also chosen to model the demand pattern as a stochastic demand pattern with Auto Regressive and Moving Average (ARMA) components of order one, i.e. ARIMA (1,0,1). This is a weakly non-stationary demand pattern with wide applicability, Box and Jenkins (1970). So studying the response to the ARMA demand pattern is particularly important as our analysis will be relevant for any echelon of a supply chain that uses the OUT policy. Furthermore, the i.i.d, AR and MA demand patterns are also instances of the general class of ARMA demands.

We also generalise the classical OUT policy in an important way. We incorporate proportional controllers into the OUT policy in order to increase the flexibility of the policy when balancing the inventory and capacity related costs. Finally, the paper also explicitly integrates the order rate variance amplification, the inventory level variance and customer service levels for 15 real life demand patterns.

The combination of the ARMA demands and the generalised Order-Up-To policy results in an (myopic and not necessarily optimal) adaptive base-stock control policy in which the base stock is adjusted as the demand forecast changes (also see Graves (1999)). It should be remembered that we are modelling a generic single echelon of a supply chain. This echelon could be a retailer, distributor or manufacturer.

Our paper proceeds as follows. Section 2 reviews literature and section 3 introduces the generalised OUT policy, which we analyse under the assumption that demand is i.i.d. This allows a concise presentation of our methodology. Section 4 investigates the response of the classical OUT policy in response to the ARMA demand pattern. Section 5 considers our generalised OUT policy in response to the ARMA demand pattern. Section 6 highlights the link between ARMA demand, the OUT policy and customer service levels for 15 real life demand patterns. Section 7 concludes.

2. Literature Background

Bullwhip has become a short-hand expression, originally coined by Procter and Gamble (Schmenner, 2001) but popularised by Lee et al (1997b), for the phenomenon where the variance of the demand signal increases as the demand signal flows up the supply chain. It has been an academic concern for a long time as initiated by Forrester (1958) and Magee (1956). There is certainly no lack of empirical evidence from industry. For example, Holmström (1997) documents a confectionary supply chain where the bullwhip increases the variance of the orders by 9 to 1 in a high volume product and 28 to 1 in a slow moving product.

Baganha and Cohen (1998) present empirical evidence and theoretical results based on the (s, S) model and highlight the link between bullwhip and inventory variance for Auto Regressive (AR) demand. Chen, Drezner, Ryan and Simchi-Levi (2000) showed the classical OUT policy with exponential smoothing and moving average forecasting for AR demands always resulted in bullwhip. Furthermore they extended these results to a multi-echelon environment both with and without information sharing. Dejonckheere, Disney, Lambrecht and Towill (2003) showed that the classical OUT policy with exponential smoothing, moving average and demand signal processing forecasting and a dynamic target "orders placed but not yet received" or WIP target will produce bullwhip for all demands, not just AR demands. They also showed how to modify the OUT policy to create a smoothing rule, i.e. one that is able to produce replenishment orders with less variance than the demand rate.

The base-stock model that we propose in this paper can be classified as an adaptive control policy. Several components adjust the order-up-to level; the forecast of demand over the lead-time, a net stock discrepancy and a pipeline stock discrepancy term. There is a rich literature on adaptive inventory control models (see e.g. Treharne and Sox (2002) and Lovejoy (1992)). These papers focus on the impact of adaptive policies on expected inventory related costs. In this contribution however we focus not only on inventory aspects, but we also investigate the bullwhip consequences. Given the complexity of the problem, we do not establish the optimality of the policy.

Sterman (1989) has presented seminal work on the effect of human behaviour on the bullwhip problem via a tabletop management game, the MIT "Beer Game". Based on an analysis of 2000 sets of results he showed that a player's decision-making behaviour could be mimicked by an anchoring and adjustment heuristic. There are also many inventory related theoretical results available, but only a small proportion of the literature relates it to the bullwhip problem; for example see Graves (1999), Deziel and Eilon (1967) and Magee (1956). We firmly believe that we should focus more on analysing inventory models not only from an inventory holding/backorder cost perspective but also on the impact these replenishment rules have on the bullwhip effect. These are the two sides of the coin. That's what we demonstrate in this paper. However, once we can describe the probability density function of the inventory levels, we can determine the customer service provided by the replenishment rule, see Deziel and Eilon (1967). Customer service can be measured in various ways, for example: the percentage of periods that inventory levels are positive at the end of the planning period, the percentage of product shipped on time to the customer, or the expected length of a stock out. We have selected a volume service level measure (fill rate, i.e. the percentage of volume supplied from shelf).

In essence, we use exactly the same methodology as the Box and Jenkins (1970) approach. However rather than use the backward difference operator we exploit the z-transform as an analysis tool. The use of transform methods to solve production and inventory control system problems has a long history. Herbert Simon initiated this research stream with an analysis in continuous time (Simon, 1952) using the Laplace transform. This approach was quickly replicated in discrete time by Vassian (1955) using the newly developed z-transform. The z-transform is the discrete time analogue of the Laplace transform. Yakov Tsypkin (1964) documents a detailed list of properties of the Discrete Laplace Transform that are easily converted into z-transform notation, many of which we exploit in this paper.

Surprisingly few publications have exploited the z-transform for discrete time analysis. Indeed continuous time approaches appear to have dominated the research field, see Axsäter (1985). However, Vassian (1955), Magee (1956), Brown (1962) and Deziel and Eilon (1967) are significant contributions that explicitly exploit exactly the same techniques as we do here using the z-transform[2].

Deziel and Eilon (1967) studied a variant of the OUT policy with a different order of events than those considered here and the analysis was conducted via computer simulation. It is this latter research that inspired this investigation. A possible cause for the relative lack of recent interest in the z-transform method is the amount of algebraic manipulation required. However, recent advancements in software (for example Mathematica by Wolfram Research, Illinois) have made the method very attractive, as there are some very powerful theorems to exploit. We will use this to derive bullwhip and inventory variance ratios for the OUT policy, provide a solution to the bullwhip problem, and study the implications of our solution on inventory levels and customer service in an extended range of demand processes.

3. A generalised Order-Up-To (OUT) policy under strictly stationary demand

In this Section we assume the demand is a stationary (i.i.d.) stochastic normally distributed random process, or white noise. White noise is any stationary stochastic process whose spectral density is a constant[3]. That is, it contains frequency components of equal amounts. A stationary i.i.d. demand process defined by (1)

$$\left.\begin{aligned} D_0 &= \mu \\ D_t &= \mu + \varepsilon_t \end{aligned}\right\} \tag{1}$$

where D_t = Demand in time *t*, μ = the mean or level of demand, ε_t = a standard normal variant at time *t*; i.e. $N(0,1)$; ε_t is a white noise stochastic process.

In an order-up-to system, the ordering decision is as follows:

$$O_t = S_t - \textit{inventory position}_t \tag{2}$$

where O_t is the ordering decision made at the end of period *t*, S_t is the order-up-to level used in period t and the inventory position equals net stock plus on order (or WIP), and net stock equals inventory on hand minus backlog. The order-up-to level is updated every period according to

$$S_t = \hat{D}_t^{\,L} + k\ \hat{\sigma}_t^{\,L} \tag{3}$$

where $\hat{D}_t^L$ is an estimate of mean demand over L (see infra) periods (we assume that $\hat{D}_t^L = L\hat{D}_t$), $\hat{\sigma}_t^L$ is an estimation of the standard deviation of the error of the forecasted demand over L periods, and k is a chosen constant to meet a desired service level or to achieve the economic stock out level (Chen et al, 2000). To simplify the analysis many authors set k equal to zero and increase the lead-time by one, for example see Dejonckheere et al, 2003 and 2004. However, we elect to set k equal to zero and increase the lead-time by a variable a (where $a \geq 0$). This results in a more general form of the OUT model as we can then consider the impact of different customer service requirements and demand signal properties on the inventory levels.

In an OUT policy we need to forecast demand over the lead-time and the review period in order to determine the order-up-to level, S. As the process is i.i.d., the best possible forecast to use every time an order is placed $(\hat{D}_t)$ is the average of all previous demands (i.e. $\overline{D}$). This we know, from the demand process assumption, is equal to μ. Hence, $\hat{D}_t = \overline{D} = \mu$.

The order of events in our replenishment system essentially follows Vassian (1955): We receive inventory and satisfy demand throughout the planning period; at the end of the planning period we observe the inventory level and place an order. Thus, even if the physical production / distribution lead-time is zero, it does not appear in the order decision until the end of the next planning period. Hence, L includes a nominal order of events delay – the review period. In other words L not only represents the physical lead-time, Tp, but also a safety lead-time (a) and a review period ($+1$). Thus we have $L=Tp+a+1$.

Finally the order-up-to policy definition is completed as follows; inventory position equals net stock (*NS*) + products on order (*WIP*). We then successively obtain:

$$\left.\begin{aligned} O_t &= (Tp + a + 1)\,\hat{D}_t - NS_t - WIP_t \\ O_t &= \hat{D}_t + (a\hat{D}_t - NS_t) + (Tp\hat{D}_t - WIP_t) \\ O_t &= \hat{D}_t + (TNS_t - NS_t) + (DWIP_t - WIP_t) \end{aligned}\right\} \tag{4}$$

where, *TNS* is the Target Net Stock position and *DWIP* is the Desired WIP position. Now we propose that the following alteration to the OUT system is made (compare (5) with (4));

$$O_t = \hat{D}_t + \frac{TNS_t - NS_t}{Tn} + \frac{DWIP_t - WIP_t}{Tw} \tag{5}$$

Setting *Tw* equal to *Tn* is particularly beneficial for system robustness as demonstrated by Disney and Towill (2002). In such a setting we will denote both parameters as *Ti*. In the classical order-up-to policy, the order quantity is a summation of the demand forecast, a net stock discrepancy (or error) term and a WIP discrepancy term, but both the net stock and WIP errors are completely taken into account. This is a full adjustment strategy. The key difference in our decision rule as expressed in (5) is that the errors are included only fractionally (Dejonckheere et al, 2003).

Hence the errors are only partially recovered during the next ordering period. These fractional adjustments are second nature to control engineers, especially when automating a process plant (Towill and Yoon, 1982). This modification is the reason why the decision rule (5) will be able to generate orders without introducing the bullwhip effect in supply chains. In fact, this is not a new concept; it actually has a long history. For example, John, Naim and Towill (1994) studied the system defined by (5) in the Laplace domain when the two feedback loop gains were allowed to be independent of each other. Sterman (1989) showed that decision makers in the beer game mimicked a policy with proportional controllers. Deziel and Eilon studied an OUT system with proportional controllers in their seminal paper in 1967. However they studied a variant of this OUT system with a slightly different order of events than we have considered here. Magee (1956) uses proportional controllers in a production ordering policy. Disney and Towill (2002) have shown that this ordering decision is guaranteed to produce a stable response if *Ti*>0.5.

As here we are studying the OUT policy reacting to i.i.d. demand here, the OUT can be further simplified to

$$O_t = \mu + \frac{(a\mu - NS_t)}{Ti} + \frac{(Tp\mu - WIP_t)}{Ti}\,. \tag{6}$$

Balakrishnan, Geunes and Pangburn (2004) propose a general (linear) order smoothing policy, where the order quantity is a convex combination of previous demand realisations. It can be shown that our smoothing rule (6) is a special case of the Balakrishnan et al. replenishment rule. Our research group and Balakrishnan et al. independently came to the same conclusion. It is interesting to see that different methodological approaches converge to the same insights. Using the procedure described in Appendix A we derive the following closed form expressions for the bullwhip (7) and inventory variance (8).

$$Bullwhip = \frac{\sigma_O^2}{\sigma_D^2} = \frac{1}{2Ti - 1} \tag{7}$$

In (7) σ_D^2 denotes the variance of demand (input), and σ_O^2 is the variance of orders (output).

Interestingly we note that under the assumptions of stationary i.i.d demand, bullwhip is independent of the lead-time. The classical OUT policy's order rate variance amplification ratio is unity. By using a *Ti*>1 the generalised OUT policy will dampen order variance amplification.

As we already mentioned in the introduction, this paper also focuses on the variance of the net stock because this has an impact on customer service. We therefore introduce the metric *NSAmp* defined by (8),

$$NSAmp = \frac{\sigma_{NS}^2}{\sigma_D^2} = 1 + Tp + \frac{(Ti-1)^2}{2Ti-1} = Tp + \left((Ti)^2 \times Bullwhip\right). \tag{8}$$

In which σ_{NS}^2 denotes the variance of the Net Stock. From (8) we learn that for stationary i.i.d demand:

- If Ti =1, i.e. a "chase sales" strategy, then *NSAmp* = 1 + *Tp* and is minimal, as we would expect as the OUT policy is optimal at reducing inventory related costs.
- If Ti >1 or Ti < 1 then *NSAmp* increases.
- *NSAmp* contains a lead-time component, $1 + Tp$, and a smoothing component, $\frac{(Ti-1)^2}{2Ti-1}$.
- Decreasing the lead-time (*Tp*) reduces *NSAmp*.
- The longer the lead-time the smaller the relative importance of *Ti* on *NSAmp* as by increasing *Ti*,
- *NSAmp* approaches *Tp*+*Ti* asymptotically.
- For a "level scheduling" strategy, $Ti = \infty$, in which case the net stock variance amplification is ∞.

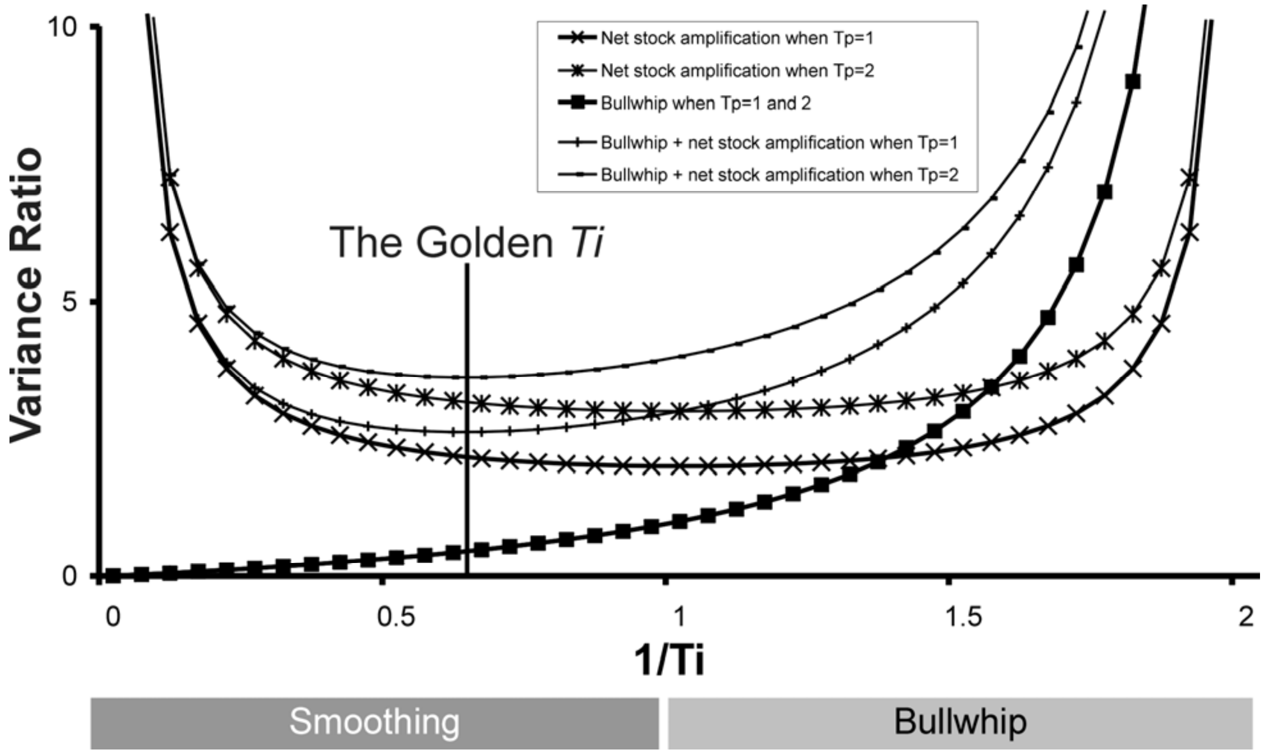

Figure 1. Bullwhip produced by the OUT policy in response to i.i.d. demand

We may plot the bullwhip and net stock amplification ratios as shown in Figure 1. Here we have plotted 1/*Ti* as it allows the complete domain to be viewed concisely. It is easy to see the role of the lead-time (*Tp*) in the net stock variance amplification ratio. Interestingly the net stock variance amplification is symmetrical about *Ti*=1, but bullwhip is not. Clearly, the proportional controllers can be used to remove bullwhip in the OUT policy. This is achieved by tuning *Ti* until the appropriate amount of smoothing is achieved, at the cost of holding extra inventory. The fact that bullwhip is not symmetrical about *Ti*=1 and that by changing *Ti* we are indeed capable of reducing bullwhip can intuitively be understood with the following analogy; "Consider for a moment the act of taking a shower. *Ti* is like a tap in a shower. The time the water takes to travel through the shower pipe is a lead-time. Just as it takes time for hot/cold water to travel through the shower to fall on our head, it takes time for an order to be produced by our factory (or delivered by our supplier). Standing in the shower we know that to get the water temperature just right, we must turn the taps very slowly. Therefore, in a supply chain we must turn the taps (*Ti*) very slowly to match supply with demand to avoid bullwhip". In this analogy, if we turn the tap at an equal speed to the demand changes (*Ti*=1) we chase demand. Turning the taps more slowly (by setting *Ti*>1) will smooth the orders and over-reacting (by setting *Ti*<1) to the demand changes will result in bullwhip.

Figure 1 illustrates the trade-off between production (replenishment) order rate and net stock variability (note that the bullwhip curve is the same for all values of *Tp*). A "best of both worlds" solution, minimising the sum of bullwhip and net stock variance amplification, is to set *Ti*=1.61803, the "Golden Ratio", for all lead-times[4]. By adding up the bullwhip effect metric and the net stock amplification metric, we assume that both factors are equally important. The "Golden Ratio" solution holds under that assumption. It is clear that in the real world companies may consider bullwhip effect related costs as more important than inventory/customer service related costs (or vice versa). In this case the shape of the "total cost curve" may be different and the optimal smoothing parameter may no longer be "golden". It is interesting, however, that the golden solution exists in the simplest scenario.

3.1. Customer service implications for i.i.d. demand

Now we will turn our attention to customer service. We have elected to use the "fill rate" as a suitable Customer Service Level (CSL) metric (Zipkin, 2000 and Silver, Pyke and Peterson, 1998). The fill-rate is popular in industry and we will use it to investigate the link between bullwhip and inventory requirements (via *a*) to meet a target CSL. The target CSL could be set to position the company's service proposition in the market place or to minimise holding and backlog costs via the economic stock-out probability. We propose here to use a predetermined fill-rate target of 99.5%. We proceed first by expressing the Target Net Stock as;

$$TNS = z \times \sigma_{NS} \tag{9}$$

where z = safety factor and σ_{NS} = standard deviation of the net stock, from (8)

$$\sigma_{NS} = \sigma_D \sqrt{1 + Tp + \frac{(Ti-1)^2}{2Ti-1}} \tag{10}$$

Now, $$Fill\ Rate = 1 - \frac{expected\ volume\ of\ backorders}{expected\ demand} \tag{11}$$

Expression (11) can be rewritten as;

$$\text{Fill Rate} = 1 - \frac{\sigma_{\text{NS}} \times L(z)}{\overline{\text{D}}} \tag{12}$$

where $L(z)$ is the expected number of units backordered per period for a safety factor, z, of the standard normal distribution.

We can determine $L(z)$ from (12). Once $L(z)$ is known, we can easily determine z using standard tables. This in turn will determine the Target Net Stock (*TNS*) to be used in (9) or equivalently, expressing *TNS* as a number of periods coverage of average demand, a:

$$TNS = a\overline{D} = z \times \sigma_{NS} \tag{13}$$

While the safety factor z is related to σ_{NS}, a represents how many periods of average demand, $\overline{D}$, are covered by *TNS*. By way of illustration of how much smoothing or bullwhip removal we can buy with inventory whilst maintaining a 99.5% service level by tuning Ti, we have developed a typical numerical example, where Tp=2, $\hat{D}_t = \overline{D} = \mu = 500$ and $\sigma_D = 100$, in Table 1.

Ti	***Bullwhip***	***NSAmp***	***a*, number of periods of inventory to ensure a 99.5% fill rate**	**TNS ($a\overline{D}$) (units)**
0.6	5	3.8	0.718	359
1	1	3	0.631	316
1.61803	0.4472	3.1708	0.644	322
2	0.3333	3.3333	0.664	332
3	0.2	3.8	0.719	360
4	0.1429	4.2857	0.773	387
6	0.0909	5.2727	0.876	438
10	0.0526	7.2631	1.061	531
20	0.0256	12.256	1.446	723

Table 1. Sample results highlighting the inventory cost of Bullwhip avoidance when *Tp*= 2

Table 1 shows that we can remove 90% of bullwhip (i.e. by setting *Ti* = 6 rather then *Ti*= 1) with a quarter of a period's extra inventory (0.876 - 0.631 = 0.245), whilst maintaining 99.5% fill rate.

4. Forecasting ARMA demand in the classical OUT policy with exponential smoothing

Consider now the case of ARMA stochastic demand. This demand is characterised by (14). We have elected to use the ARMA demand pattern in order to create a situation where the use of a forecasting mechanism in the OUT policy is justified to investigate its impact on dynamic performance. ARMA is weakly stationary and for particular settings it does exhibit some non-stationary properties that can be forecasted. We note that truly non-stationary demand patterns have no natural mean and infinite variance[5]. The mean centred ARMA demand pattern can be generated from stationary white noise as follows;

$$\left.\begin{aligned} D_{\underset{ARMA}{0}} &= \varepsilon_0 + \mu \\ D_{\underset{ARMA}{t}} &= \rho\left(D_{\underset{ARMA}{t-1}} - \mu\right) - (1-\alpha)\varepsilon_{t-1} + \varepsilon_t + \mu \end{aligned}\right\} \tag{14}$$

where; ε_t = white noise, μ = mean of the ARMA demand pattern, ρ = auto regressive coefficient, -1 < ρ < 1, α = moving average coefficient, 0 ≤ α ≤ 2 and $D_{\underset{ARMA}{t}}$ = ARMA demand at time *t*. A positive autoregressive coefficient will result in "meandering demand patterns”, whereas a negatively correlated demand patterns will be more erratic over time, alternating about the mean.

Appendix A details our analysis of (14) in order to determine the variance of the ARMA demand. This procedure has been used throughout our paper to produce variance ratios, but for brevity we have not given any other examples.

Recall that the “classical” OUT policy is defined by

$$O_t = (Tp + a + 1)\,\hat{D}_t - NS_t - WIP_t \tag{15}$$

and that the policy requires an estimate or forecast of demand over the lead-time. For stationary uncorrelated demands, the best forecast of demand in the future is well known to be the average of all previous demands $\overline{D}$. However, for correlated demands such as AR and ARMA demands, a forecast ($\hat{D}$) can be produced with less forecast error than $\overline{D}$ by using a forecasting mechanism such as exponential smoothing, Muth (1960). This is defined in (16) where *Ta* is the average age of the demand data in the forecast. *Ta*>-0.5 ensures a stable response, because the range $Ta \in (-0.5, \infty]$ corresponds to $\beta \in (0,2]$ as $\beta = \frac{1}{1+Ta}$ so,

$$\hat{D}_t = \hat{D}_{t-1} + \frac{1}{1+Ta}\left(D_t - \hat{D}_{t-1}\right). \tag{16}$$

We have selected exponential smoothing as it is well understood and popular with practitioners. For example, empirical research by Makridakis et al. (1982) has shown simple exponential smoothing to be a good choice for one-period-ahead forecasting. It was the preferred option from among 24 other commonly used time series methods compared under a variety of accuracy measures and theoretical models for the process underlying the observed time series.

As shown in Appendix B, we may investigate the performance of exponential smoothing in response to the ARMA demand and determine the optimum smoothing parameter, *Ta* that will minimise the one period ahead mean squared forecast error for particular values of α and ρ. The resulting closed form for the optimal *Ta* is given by (17) which we have plotted in Figure 2 for various α and ρ.

$$Opt\_Ta = \frac{\begin{pmatrix} (\alpha-2)^2 - 6\rho(1-\alpha) - 2\alpha^2\rho + 2\rho^2(1-\alpha) - \\ (\alpha-2)\sqrt{\rho(-1+\alpha+\rho)(1+(-1+\alpha)\rho)} \end{pmatrix}}{-4(-1+\rho)^2 + 4\alpha(-1+\rho)^2 + \alpha^2(-1+3\rho)} \tag{17}$$

(17) results in negative or complex values recommendations if $\rho < 1 + \frac{\alpha\left(-3\alpha + \sqrt{32 + \alpha(-32+9\alpha)}\right)}{8(-1+\alpha)}$.

This expression is very nearly, but not quite, $\rho < 1 - \frac{2\alpha}{3}$ for α<1. In this case *Ta*=∞ should be used, as exponential smoothing will not produce a forecast with less mean squared error than the unconditional mean of the demand process, μ. It should be remembered that our recommended *Ta* is optimal for minimising the one period ahead forecast error and we have defined the Order-Up-To level as $S = (1 + a + Tp)\hat{D}_t$ in this analysis. We do not claim *Ta* to be optimal at minimising inventory / shortage or bullwhip (or their sum) costs or that this is the optimal way of calculating *S* or that *OptTa* minimises the forecast error of the demand over the lead-time.

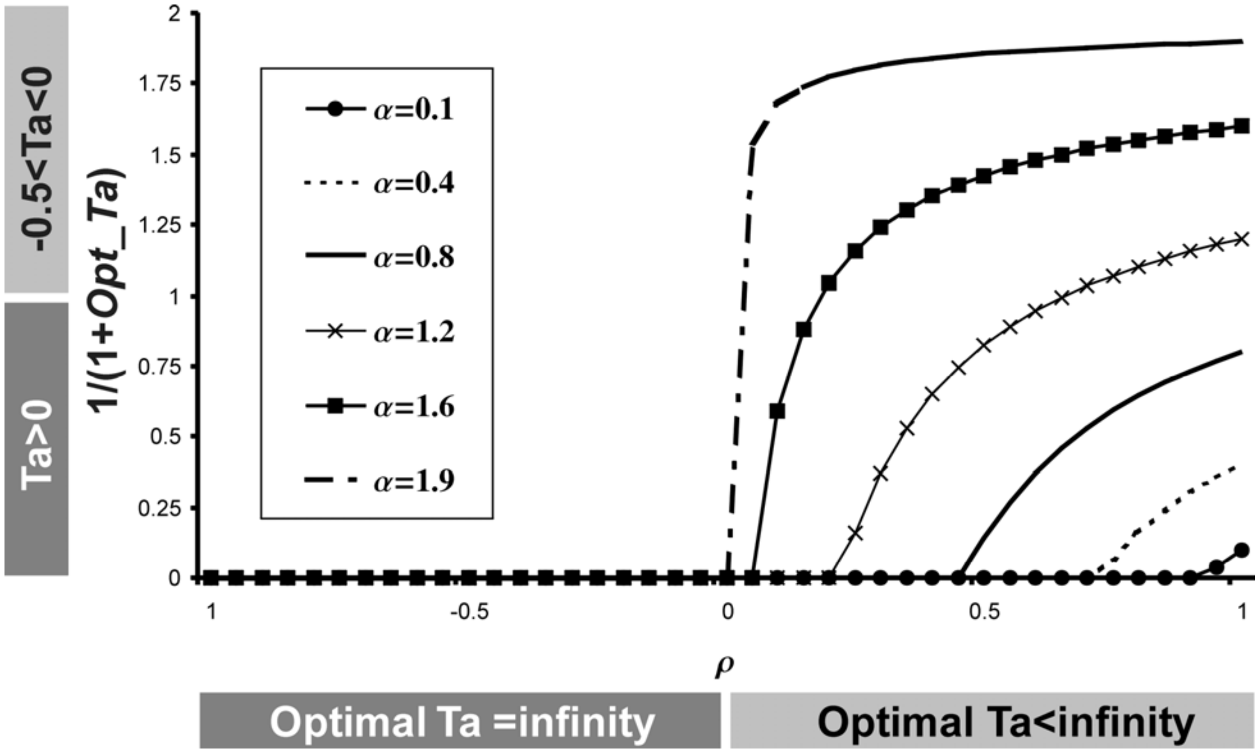

Figure 2. The optimal exponential smoothing forecasting parameter (*Ta*) that minimises the ARMA one period ahead forecasting error

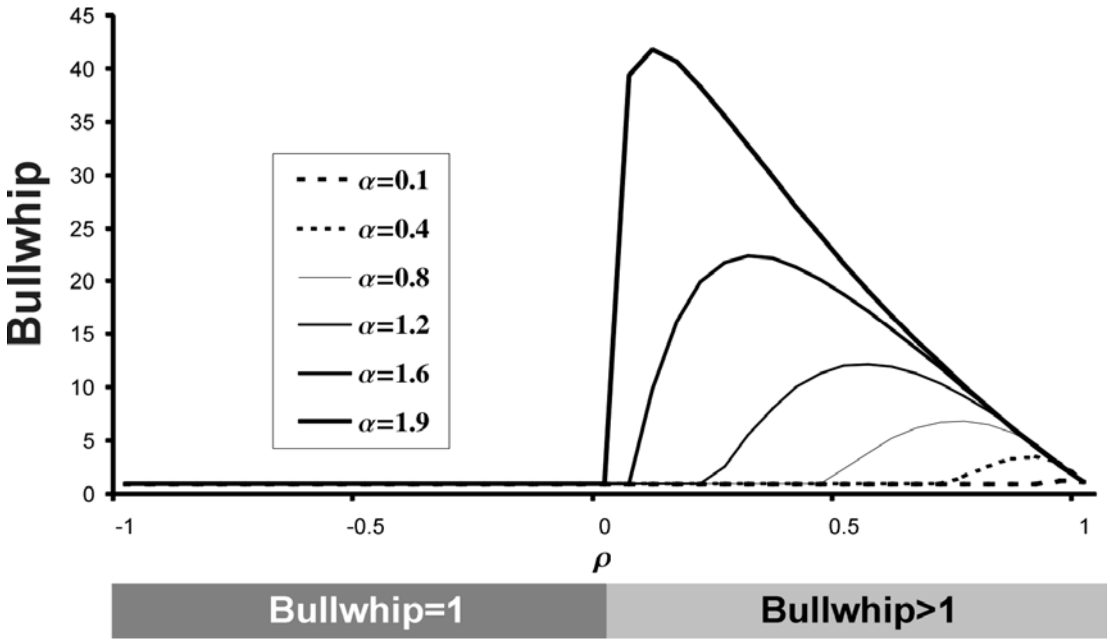

Figure 3. Enumeration of (B3) with *Tp*=2, *a* =1 and an optimal *Ta*

Recall that the classical OUT system simply passes on orders for i.i.d. demands. Furthermore, an exponential smoothing forecasting mechanism will always produce a forecast with less variance than the ARMA demand. This can be determined from (B1), a closed form expression for the forecasting variance amplification ratio. We ask ourselves… "So why is there a bullwhip problem?" If our forecast has less variance than demand, why can't our orders have less variance than demand? The answer is that it is the combination of the forecasting mechanism, order delay and inventory feedback loops in the OUT system that causes bullwhip. Remember that in the "classical" OUT policy, the order quantity is the summation of the demand forecast and two inventory feedbacks, the net stock discrepancy term and the WIP discrepancy term (see equation (4)). The weights attached to these discrepancies are each time one (the so-called full adjustment strategy). This full adjustment strategy causes bullwhip. The key difference with the generalised OUT policy is that the errors in the feedback loops are only included fractionally (see section 5). This smoothing principle is also discussed in Balakrishnan, Geunes and Pangburn (2004).

Using our methodology, the closed form expression (B3) shown in Appendix B is obtained for the bullwhip produced by the "classical" OUT policy when *Ti*=1. It can be shown that (B3) is always greater (or equal to) than one for all α and ρ. We have plotted the bullwhip produced in response to some ARMA demands in Figure 3. For each case, *Ta* was set as defined by (17). We can see that the "classical" OUT policy with *Ti*=1 produces bullwhip when exponential smoothing forecasting is used. In fact, it is known that the "classical" OUT policy with exponential smoothing forecasting produces bullwhip for **all** demands from a frequency domain analysis (Dejonckheere, et. al. 2003), but we can now also confirm this for ARMA demands. Note that in Figure 3 there is a zone where the bullwhip effect equals 1. This happens for negatively correlated demands, and the optimal *Ta* is infinity in this case; consequently the exponential forecasting is replaced by the long term average demand (see expression (16)). This is the case of the "classical" OUT policy, using $\overline{D}$ as a forecaster, where it is well known that the bullwhip effect equals one.

To summarise this section we have shown how to tune the exponential smoothing to minimise the one period ahead forecast error in response to ARMA demand. We note that the unity gain in the two feedback loops induces bullwhip in the classical OUT policy with exponential smoothing forecasting.

5. The generalised OUT policy under ARMA demand

Now we consider the case of the generalised OUT with an exponential smoothing forecasting mechanism in response to ARMA demand. Recall, our generalised OUT policy:

$$O_t = \hat{D}_t + \frac{TNS_t - NS_t}{Ti} + \frac{DWIP_t - WIP_t}{Ti} \tag{18}$$

Note, (18) is exactly the same as the "classical" policy studied in section 4 except for the proportional controller ($1/Ti$). The bullwhip and net stock amplification ratios for this generalised OUT policy in the complete ARMA plane are tractable but they are very lengthy and are shown in Appendix B. They actually have a number of very nice properties. Firstly, as all the ARMA demands when $\alpha + \rho = 1$ are i.i.d., the variance ratios are the same as those presented in Section 3.

For ARMA demands when $\alpha + \rho$ is only slightly greater than 1, the characteristic U-shaped inventory variance curves flexes to the right. See Figure 4 and Figure 6 where the average inventory holding (*a*) is one period of demand. It is better in terms of inventory holding and backlog costs to use a Ti<1, i.e. lower inventory variability is achieved by over-reacting to the

ARMA demand signal. This is intuitive as we are effectively gambling on trends in demand having a lasting impact, and over-reacting to changes in demand will reduce the error between demand and supply after the lead-time, thus reducing inventory requirements. Hence in these situations if we want to remove bullwhip, we will be forced to hold extra inventory (when compared to case when *Ti* is set to minimise inventory costs and when compared to the case of the "classical" OUT policy).

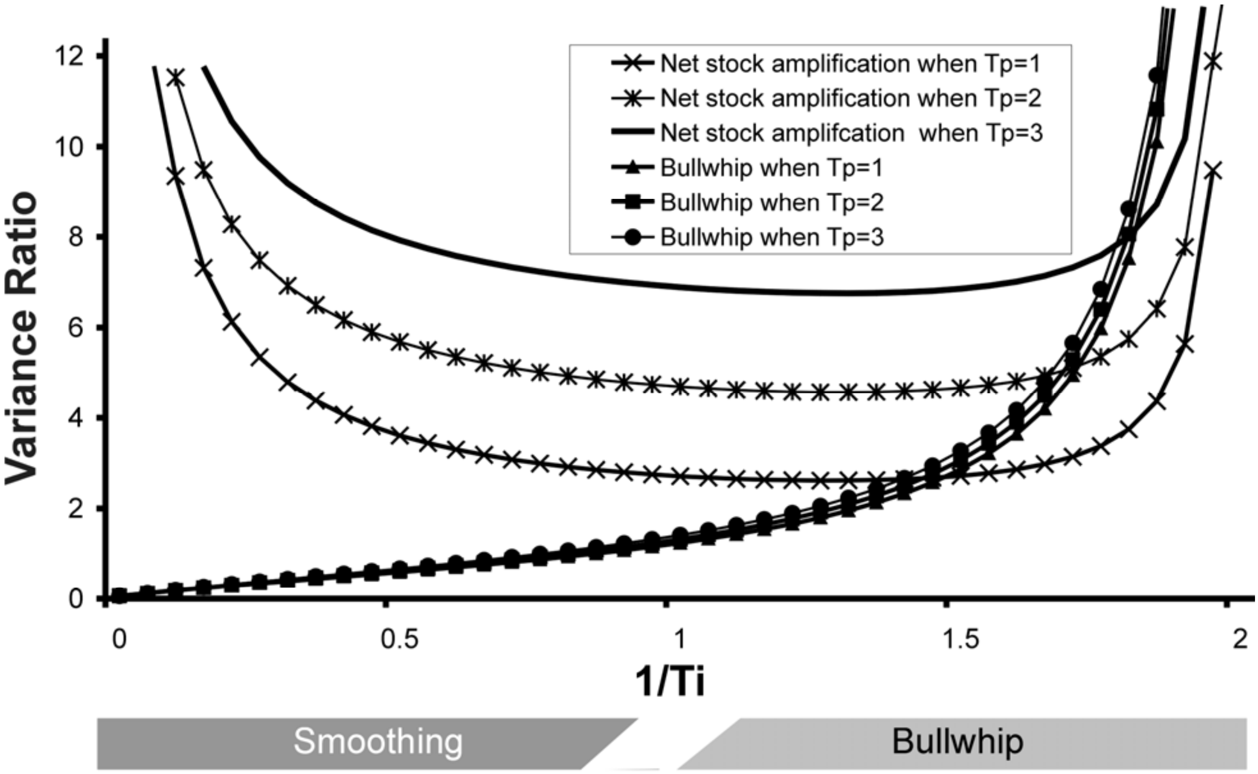

Figure 4. Bullwhip and net stock amplification when α =0.75 and ρ =0.5, *Ta* = 25.22

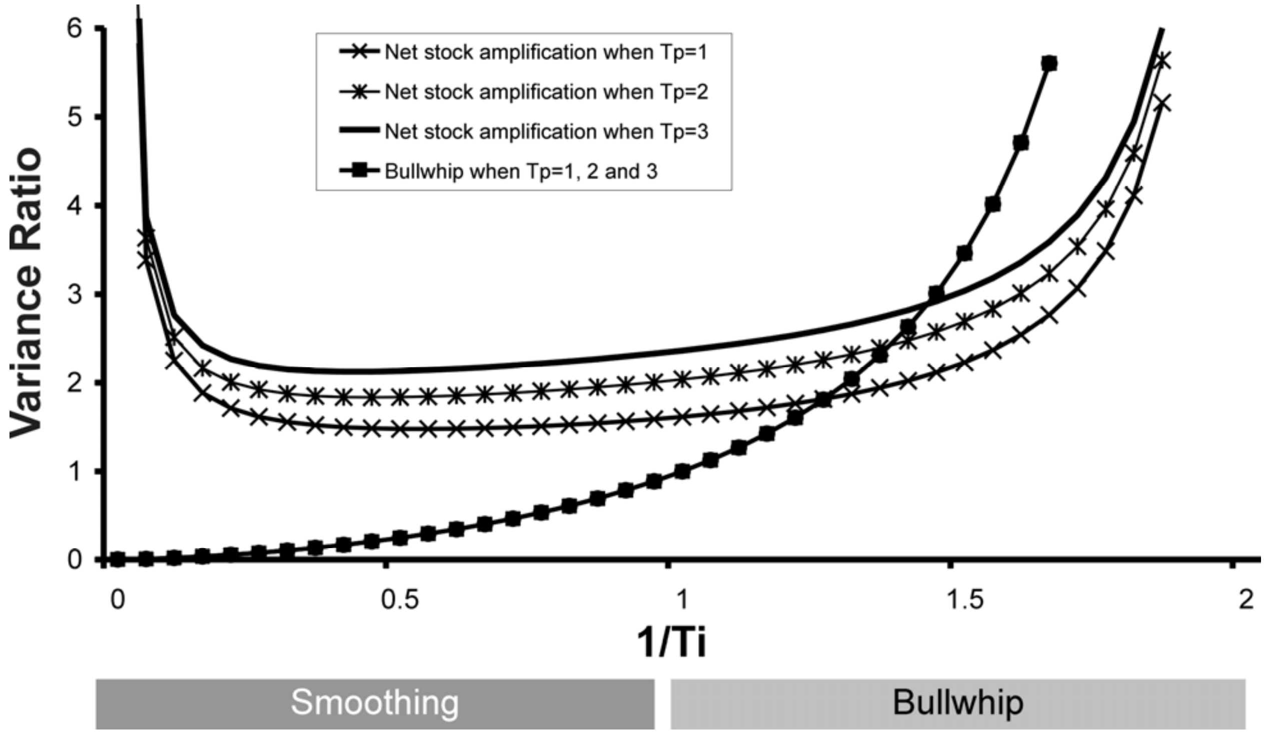

Figure 5. Bullwhip and net stock amplification when α=0.25 and ρ =0.5, *Ta* =∞

However, if $\alpha + \rho < 1$ and when $\alpha + \rho$ is much greater than 1, the U shaped net stock curve flexes to the left (see Figure 5, where *Ta* has been set to minimise the one-period ahead forecast error and Figure 6). Inventory variability is reduced by smoothing the demand signal (*Ti*>1). In this case, bullwhip can be removed whilst reducing net stock variance (when compared to the "classical" OUT policy at *Ti*=1). We can also see in Figures 4 and 5 the role of the lead-time, *Tp* and the feedback gain, *Ti* on bullwhip and inventory variance. Lead-time (*Tp*) increases the net stock variance for all cases but its effect is greatly reduced when $\alpha + \rho$<1. For ARMA demands when $\alpha + \rho$>1, *Tp* also increases bullwhip[6], something that did not happen for i.i.d demands or for the ARMA demands when $\alpha + \rho$<1. This effect has been introduced by the forecasting mechanism. *Ti*'s symmetrical impact on net stock variance and bullwhip curves has been influenced by the ARMA constants.

The precise manner in which this curve bends to the right or to the left is described in Figure 6. We can see that sometimes the inventory variance curve bends to the right (region B in Figure 6), in which case, if we want to avoid bullwhip then the customer service level achieved with the one period inventory holding decreases when compared to the "classical" OUT policy. When the inventory variance curve flexes to the left (regions A and C in Figure 6), bullwhip reductions may be achieved whilst simultaneously improving the customer service levels offered by the one period's inventory holding, when compared to the "classical" OUT policy. We notice that most of the ARMA demands result in this win-win scenario.

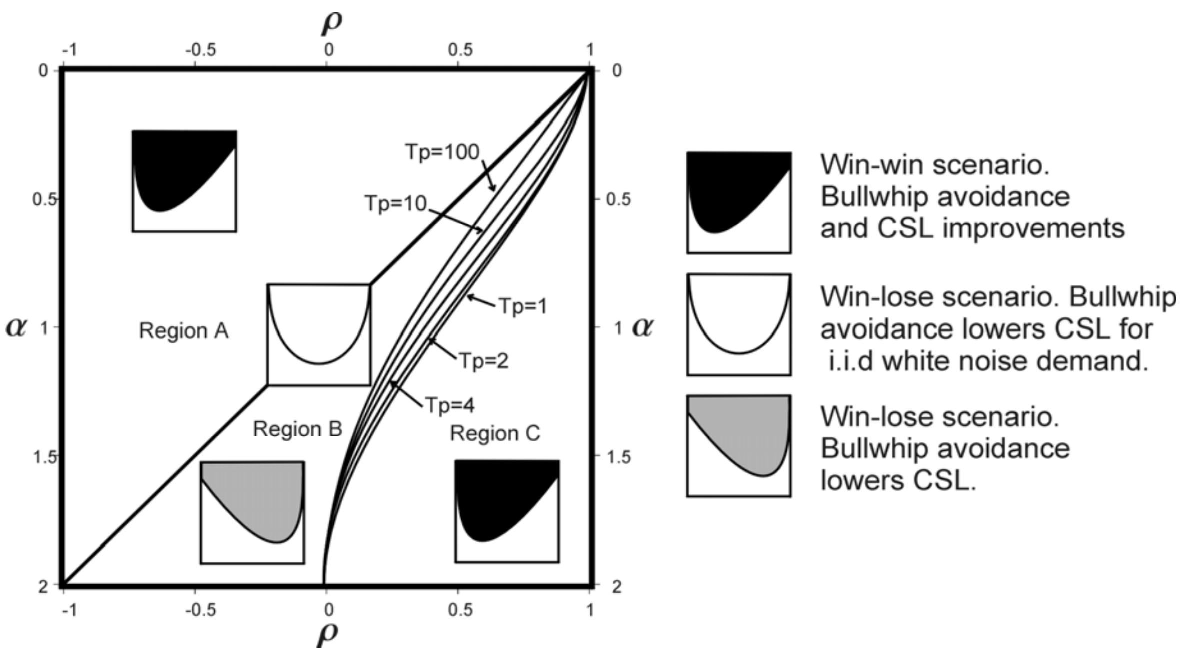

Figure 6. Net stock variance behaviour of the generalised OUT policy with optimal one period ahead forecasting in the ARMA plane for various lead-times when *a*=1

For situations where the optimal $Ta = \infty$, the complex bullwhip and inventory variance expressions ((B4) and (B5)) simplify a little. They are shown in Table 2 for different classes of the ARMA demand pattern. We can see clearly here that when we use the unconditional mean as the forecast in the OUT policy that bullwhip is independent of lead-time and reducing the lead-time reduces *NSAmp*. Furthermore ρ has a smaller relative impact on *NSAmp* for longer lead-times. We can also confirm the results of Chen et al (2000) for AR demands, that is; positively correlated demands decrease bullwhip and negatively correlated demands increase bullwhip in the OUT policy with exponential smoothing forecasting.

Demand Pattern	**Bullwhip**	**NSAmp**
i.i.d.	$\frac{1}{2Ti-1}$	$Tp+1+\frac{Ti^2-1}{2Ti-1}=Tp+\frac{Ti^2}{2Ti-1}$
AR(1)	$\frac{1}{(2Ti-1)}\frac{(Ti(1+\rho)-\rho)}{(Ti(1-\rho)+\rho)}$	$\frac{\left(\frac{(Ti^2+Tp(2Ti-1))(Ti(1+\rho)-\rho)}{2Ti-1}+\frac{2\rho(Tp(1-\rho)-\rho(1-\rho^{Tp}))}{(1-\rho)^2}\right)}{Ti(1-\rho)+\rho}$
MA(1)	$\frac{1}{(2Ti-1)}\frac{2(1-\alpha)+Ti\alpha^2}{Ti\left(1+(1-\alpha)^2\right)}$	$\frac{1}{(2Ti-1)}\frac{2Ti(1-\alpha)+\left(Ti^2+Tp(2Ti-1)\right)\alpha^2}{\left(1+(1-\alpha)^2\right)}$
ARMA (1,1)	$\frac{1}{(2Ti-1)}\frac{\left(2(1-\rho)(1-\alpha)+(Ti(1+\rho)-\rho)\alpha^2\right)}{\begin{pmatrix}(Ti(1-\rho)+\rho)\\ \left(2(1-\rho)(1-\alpha)+\alpha^2\right)\end{pmatrix}}$	$\frac{\begin{pmatrix}Ti^3\alpha^2(\rho^2-1)(1-\rho)+\\ Ti\begin{pmatrix}Tp\alpha^2(\rho^2-1)(-1+3\rho)+\\ 4\rho(\alpha+\rho-1)(1-\rho+\alpha\rho)(1-\rho^{Tp})\end{pmatrix}-\\ Ti^2(1-\rho)^2(2-2\rho+\alpha(-2+2\rho-\alpha\rho+2Tp\alpha(1+\rho)))+\\ \rho\left(2(1-\alpha)(1-\rho)^2(1-\rho^{Tp})-\alpha^2\left(Tp(\rho^2-1)+2\rho(1-\rho^{Tp})\right)\right)\end{pmatrix}}{\begin{pmatrix}(2Ti-1)(Ti(-1+\rho)-\rho)(1-\rho)^2\\ \left(2(1-\rho)(1-\alpha)+\alpha^2\right)\end{pmatrix}}$

Table 2. Simplified closed form expressions for *Bullwhip* and *NSAmp* when *Ta*=∞

6. Customer service insights

We now investigate the generalised OUT policy more explicitly in terms of the Customer Service metric, the "fill-rate". We already know the variance (see (B3), (B4) and (B5) in Appendix B) and mean of the orders ($\overline{D}$) and inventory levels ($\overline{D}\,a$), hence we have all the information needed to describe the ability of our OUT policy to meet a desired fill-rate.

It is difficult to consider the complete solution space herein as it requires manipulation of the probability density function of the normal distribution that is essentially non-algebraic. So we will take two approaches to study the CSL implications. First we consider graphically the relationship between the feedback gain, *Ti*, and the inventory required to achieve 99.5% fill-rate at various points in the ARMA plane. We then analyse 15 representative demand patterns from Procter and Gamble's home-care and family-care product range with a numerical approach.

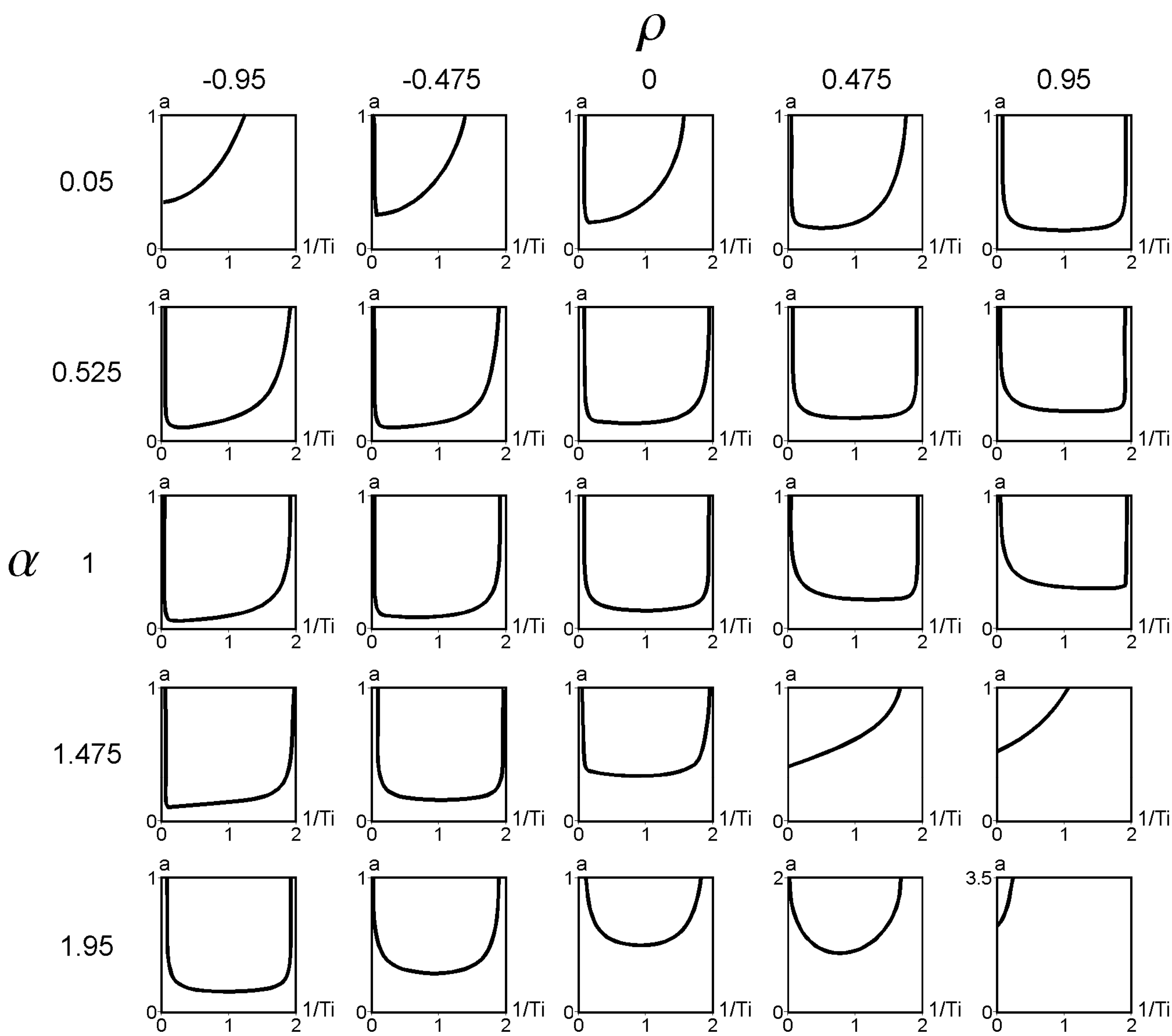

Figure 7. Average number of periods inventory holding (*a*) required to the 99.5% fill-rate objective as a function of *Ti* in the ARMA plane with one period ahead forecasting

Figure 7 details the relationship between *Ti* and *a* needed to achieve the fill-rate objective. Here *Tp*=2 and *Ta* was set to minimise the one period ahead forecast error. The contour in each plot indicates the minimum *a* required the meet the fill-rate objective. The area below the contour results in a service level below the target; with $\alpha > 1$, it becomes increasingly difficult to achieve the CSL target as ρ increases. We can see that it is possible to end up in four different scenarios when compared to the classical OUT policy (Ti=1) whilst maintaining the fill-rate objective; ***Win-Win***, we can remove bullwhip (by using a large enough Ti) and reduce inventory levels, ***Win-Lose***, sometimes bullwhip can only be removed at the expense of holding extra inventory, ***Lose-Win***, sometimes bullwhip can be endured because it results in a policy that requires less inventory to be held, ***Lose-Lose***, sometimes excessive bullwhip and inventory may exist.

We now turn our attention to the real-life demand patterns from Procter and Gamble. We identified the ARMA constants that minimised the mean squared error between the auto-correlogram of 15 real life demand patterns and the impulse response of the ARMA demand pattern. Our results are shown in Figure 8. Our co-authors from Procter & Gamble have carefully examined and exploited the impact of the OUT policies (classical and generalised) on performance in terms of inventory /customer service related costs and flexibility adjustment costs in their business.

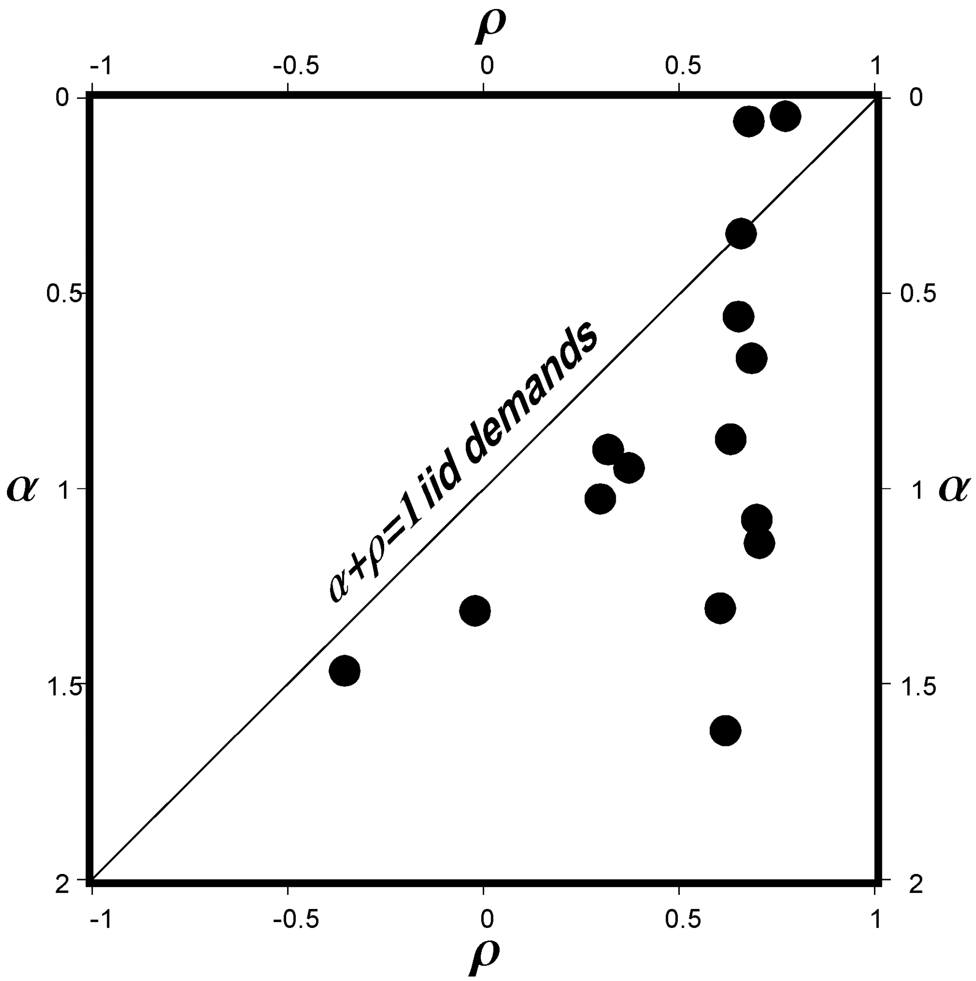

Figure 8. Real demand patterns in the ARMA parameter plane

We can see that our real life demand patterns, selected on their basis of being qualitatively similar to ARMA demand patterns, are predominantly positively correlated and lie to the right of the stationary and i.i.d. demand patterns on the $\alpha + \rho = 1$ line. However, they lie in all three regions of Figure 6.

Using the fill-rate procedure (Eqs 9 to 12) and the variance ratio expressions (B3, B4 and B5) we have investigated the link between bullwhip, inventory variance and customer service levels as follows; For *Tp*=2 and *Ta* set with (17), we then set the Target Net Stock gain (*a*) required to achieve 99.5% fill-rate and note the value of bullwhip at *Ti*=1. This serves a benchmark for the amount of inventory required to achieve 99.5% fill-rate and the amount bullwhip produced by the classical OUT policy. Next we find the value of *Ti* that will minimise *a* to achieve the CSL objective and note the value of bullwhip at this point. This gives us a setting for the generalised OUT that will minimise the average inventory holding. Our results are shown in Table 3.

α	ρ	**Ta**	**Classical OUT policy, *Ti*=1**		**Modified OUT policy**		
			Periods inventory	**Bullwhip**	***Ti***	**Periods inventory**	**Bullwhip**
0.926	0.371	∞	0.218	1	0.7322	0.2125	1.7314
1.454	-0.35	∞	0.1705	1	0.9246	0.1703	1.1580
1.133	0.711	0.041	0.498	7.9232	2.3697	0.4735	3.4673
1.024	0.289	∞	0.218	1	0.7318	0.2128	1.7128
1.072	0.694	0.149	0.465	7.7231	2.3981	0.445	3.3616
1.597	0.611	-0.325	0.725	13.228	1000	0.534	1.1841
1.296	0.607	-0.075	0.552	10.606	1000	0.446	1.0497
0.001	0.704	∞	0.143	1	400	0.1195	0.00001
0.332	0.657	∞	0.1559	1	1.0251	0.1558	0.9516
0.893	0.324	∞	0.199	1	0.7855	0.1958	1.5573
1.295	-0.018	∞	0.201	1	0.7849	0.1987	1.5074
0.872	0.629	0.896	0.3505	5.6324	1.2453	0.3486	4.3868
0.658	0.673	2.383	0.2744	3.3732	0.9443	0.2741	3.6493
0.541	0.641	23.39	0.206	1.2748	0.8084	0.2029	1.8698
0.001	0.760	∞	0.145	1	64.52	0.1346	0.0005
		Average	0.3014	3.8507	Average	0.2749	1.8391
				Average percentage gain		8.77	52.23

Table 3. Numerical results for 15 real world demand patterns

It can be seen that *Ti* always allows us to reduce inventory requirements to meet the CSL when compared to the classical OUT policy (or at least match the classical OUT policy's performance). In some cases this is realised by actually reducing bullwhip, but in others, bullwhip has been increased. This yields two insights. Bullwhip avoidance, inventory reduction and CSL objectives can sometimes all be achieved simultaneously. Whilst in other situations bullwhip is actually useful, as it allows an increase in CSL and/or a reduction in inventory holding. We also note from Table 3 that *Ta* is sometimes a small negative value. In such cases the exponential smoothing forecast over-reacts to changes in the demand pattern. However, a stable forecast is achieved when *Ta*>-0.5, so these cases are perfectly reasonable recommendations.

To summarise this section we have studied the service implications of the generalised OUT policy. This was done in the general ARMA case and also for 15 real world demand patterns. Of particular interest is the conclusion that for the OUT policy, when compared to the case when Ti=1, the following four general solutions are possible; Win-Win, Win-Lose, Lose-Win and Lose-Lose when considering bullwhip and stock levels that meet the CSL objective.

It is clear that it is worth monitoring the demand statistics to determine the ARMA parameters and thereby find "better" policy settings that lead to competitive advantage. We have shown how the OUT policy can be "tuned" to suit a variety of objectives. The one that will be the best in a given situation will depend on a number of factors. For example in an industry with high inventory related costs, it may be advantageous to flex capacity. A retailer may want to reduce inventory in order to be able to offer a broad product range through its facilities. Whereas for a manufacturer, buy-backs and obsolescence may be the more significant inventory related costs. In contrast, in an industry with long production runs and high capacity related costs, exploiting inventory holding to avoid bullwhip related costs may be more economically desirable. Bullwhip related costs in a retailer may be concerned with distribution activities, whereas for a manufacturer they may result from production matters.

Clearly, a properly defined OUT policy can help industry to exploit properties of the demand signal to balance bullwhip and inventory issues or reduce them both concurrently. However, in general, there will only be a win-win scenario for certain demand patterns. We do not think one can identify upfront the likelihood of being in a win-lose or win-win scenario in a particular business without some investigation into the business's demand streams. However if demand can be characterised by the ARMA model we may use Figures 6 and 7 to gain some insight into this question.

7. Conclusions

We have shown that the adaptation of the classical OUT policy to include proportional controllers in the order decision is highly desirable. It gives considerably more flexibility to "tune" the purchasing/production/distribution ordering decision to exploit supply chain characteristics. For example, by minimising inventory exposure or by using inventory to "buy stability". How this will be achieved in a given situation clearly depends on the peculiarities of each individual case.

However it is clear we can tune an ordering system to exploit the statistical properties of the demand process. We have shown that it is possible to actually achieve bullwhip and inventory level reduction together whilst maintaining CSL. This is a true win-win situation resulting from our generalised OUT policy. However this cannot be achieved in all cases as it depends on the demand pattern. Neither is bullwhip avoidance always desirable, for instance we may even choose to induce bullwhip as it sometimes enables further inventory reductions to be achieved.

Appendix A. Deriving closed form variance amplification expressions

Our procedure for determining the closed form expressions of the variance amplification ratios will now be illustrated by example. We have chosen to use the ARMA demand variance amplification ratio as it is concise, but the procedure is essentially the same for all the ratios. Departing from the difference equation representation (14) of the demand pattern we first convert it into a z-transform model via the block diagram shown in Figure A1. For a useful overview of block diagrams and control theory we refer readers to Nise (1995).

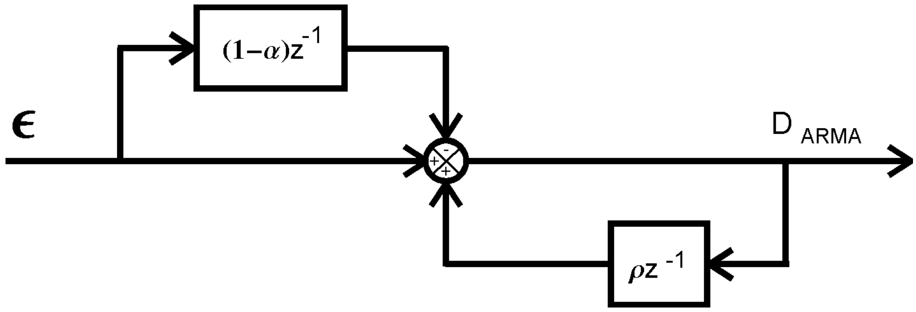

Figure A1. Block diagram of the ARMA demand generator

Manipulation of this block diagram using standard techniques yields the following transfer function (where z is the z-transform operator) that describes completely the demand pattern in the discrete complex frequency domain;

$$\frac{D_{ARMA}}{\varepsilon} = \frac{z+\alpha-1}{z-\rho} \tag{A1}$$

In order to calculate the variance amplification ratio between the pure white noise input and the ARMA demand pattern we exploit Tsypkin's relation (Tsypkin, 1964) that states that the variance of a system's output divided by the variance of the input (when the input is pure white noise) is equal to the sum of the squared impulse response in the time domain. This relationship is described by (A2) and is covered in more detail in Disney and Towill (2003). A summary of Tsypkin's relationships between; the variance amplification ratio (VR) in question, its statistical definition, the system transfer function via the area under the squared frequency response, the noise bandwidth, W_N, and sum of the squared impulse response in the time domain, is as follows;

$$VR = \frac{\sigma^2_{Output}}{\sigma^2_{Input}} = \frac{1}{\pi}\int_0^{\pi}\left|F(w\sqrt{-1})\right|^2 dw = \frac{W_N}{\pi} = \sum_{n=0}^{\infty} f^2[n] \tag{A2}$$

The ARMA lattice function ($f[n]$) is the inverse z-transform of (A1) and is given by;

$$f_{D_{ARMA}}[n] = \rho^{n-1}\left(\rho + (\alpha - 1)h[n-1]\right) \tag{A3}$$

where $h[x] = \begin{cases} 0, x < 0 \\ 1, otherwise \end{cases}$ is the Heaviside step function. Using (A2), the variance ratio of the ARMA demand is therefore given by (A4). We have used this technique throughout this paper, without referring to or presenting the details, as the equations involved are often very lengthy.

$$ARMAAmp = \sum_{n=0}^{\infty}\left(\rho^{-1+n}\left(\rho + (\alpha - 1)h[n-1]\right)\right)^2 = \frac{2 - 2\rho + \alpha(\alpha + 2\rho - 2)}{1-\rho^2} = 1 + \frac{(1-\alpha-\rho)^2}{1-\rho^2} \tag{A4}$$

Appendix B. The closed form ARMA forecasting and OUT variance ratios

For brevity, only the starting point in the analysis (the block diagram) and the end point (the variance amplification ratio) are presented. Our results were obtained by using Mathematica (Wolfram Research, Illinois) to assist in the algebraic manipulation required. They were crosschecked by comparison to a difference equations simulation in a spreadsheet and via previous results in the literature where available (for example Chen et al 2000).

B.1 The ARMA forecast variance ratio

The required block diagram is shown below. Note that $\beta = 1/(1+Ta)$. The procedure is as follows; Manipulate the block diagram for the $\hat{D}_{ARMA}/\varepsilon$ transfer function, take the inverse z-transform, exploit Tsypkin's relation (A2) to find the variance ratio ($\sigma^2_{\hat{D}_{ARMA}}/\sigma^2_{\varepsilon}$) and finally divide by the variance of D_{ARMA}. This yields (B1), the variance amplification ratio of exponential smoothing forecasting mechanism when reacting to ARMA demands.

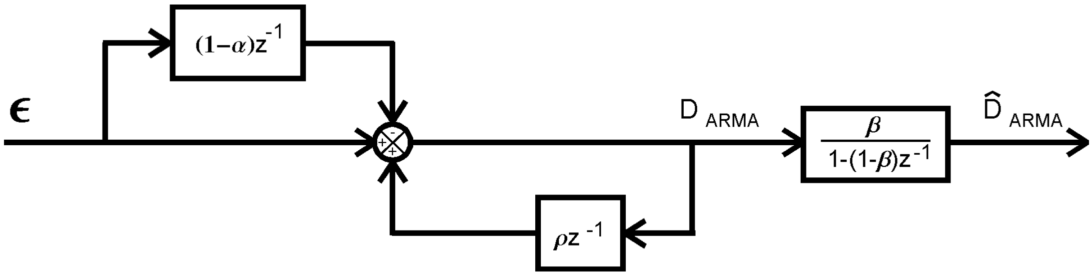

Figure B1. Block diagram of the exponential forecast of ARMA demand

$$FAmp=\frac{\sigma^2_{\hat{D}_{ARMA}}}{\sigma^2_{D_{ARMA}}}=\frac{2\alpha-\alpha^2-Ta\alpha^2+2\rho-2\alpha\rho-Ta\alpha^2\rho-2}{(1+2Ta)(Ta\rho-1-Ta)(2-2\alpha+\alpha^2-2\rho+2\alpha\rho)} \tag{B1}$$

B.2 The variance of one period ahead forecast error

Minimising the variance of the one period ahead forecast error is the same as minimising the mean squared error of the one period ahead forecast error. Thus we may differentiate (and solve for zero gradient) the variance of the forecast error to derive the optimal *Ta* for forecasting ARMA demand. It is not necessary to divide by the ARMA variance. The necessary procedure to find the variance of the forecast error is to; find the z-transform of the block diagram, take inverse and sum its square from zero to ∞ to give (B2).

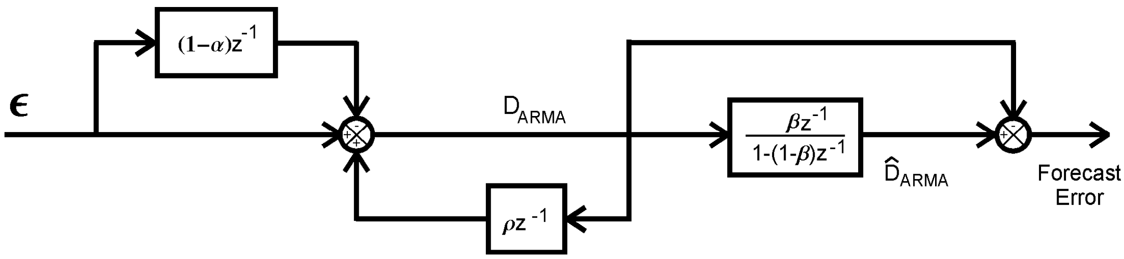

Figure B2. Block diagram of the forecast error transfer function

$$\frac{\sigma^2_{FE}}{\sigma^2_{\varepsilon}}=-\frac{2(1+Ta)(3-\rho+\alpha(\alpha+\rho-3)+Ta(2-2\rho+\alpha(\alpha+2\rho-2)))}{(1+2Ta)(Ta(\rho-1)-1)(1+\rho)} \tag{B2}$$

B.3 Order rate variance ratio

We have used three bullwhip expressions in our story. They were all found using the following block diagram. In Section 3 $\beta = 0$, Section 4 $\beta = \frac{1}{1+Ta}$, Ti=1 and in Section 5 $\beta = \frac{1}{1+Ta}$.

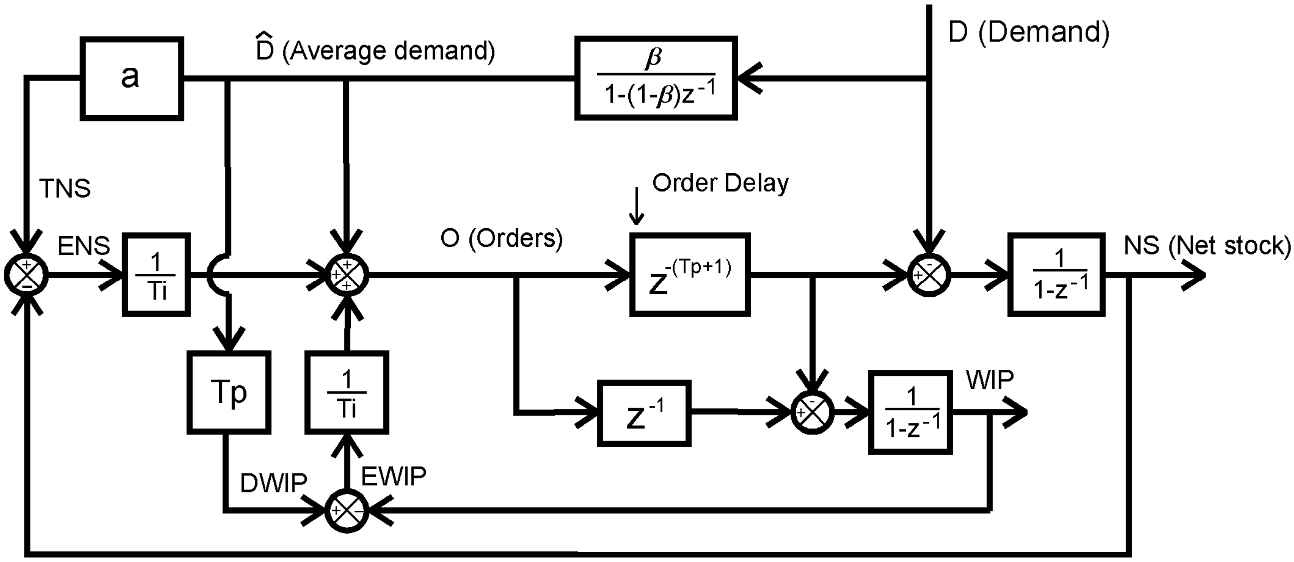

Figure B3. Block diagram of the generalised OUT policy

The procedure to derive the bullwhip expression was the same in all three cases, that is; re-arrange the block diagram to get the order rate transfer function of the OUT policy, multiply by the transfer function of the demand signal, take the inverse z-transform of the product, find the closed form sum of the square of the difference equation and divide the result by the variance of the demand pattern.

The closed form bullwhip equation for the generalised OUT policy reacting to i.i.d demands was shown in (7). The bullwhip expression for the classical and generalised OUT policy in response to ARMA demands is given by (B3) and (B4) respectively, where $T\overline{p} = Tp + a$, $\kappa = 2 - 2\rho + \alpha(\alpha - 2 + 2\rho)$, and $\lambda = 3 - \rho + \alpha(\alpha + \rho - 3)$.

$$\frac{\sigma^2_{O,cOUT}}{\sigma^2_{D_{ARMA}}}=\frac{\begin{pmatrix}14\alpha-14-5\alpha^2+2(-3+\alpha)\rho+4(-1+\alpha)\rho^2+\\ Ta\begin{pmatrix}4(7+T\overline{p}(6+T\overline{p}))\alpha-4(7+T\overline{p}(6+T\overline{p}))-2(2+T\overline{p})(3+T\overline{p})\alpha^2+\\ \left(42+10T\overline{p}(\alpha-2)^2+2T\overline{p}^2(\alpha-2)^2-42\alpha+9\alpha^2\right)\rho+2(7+2T\overline{p}(4+T\overline{p}))(\alpha-1)\rho^2\end{pmatrix}+\\ 6T\overline{p}(\rho-1)\lambda+2T\overline{p}^2(\rho-1)\lambda+2Ta^3(\rho-1)\kappa+Ta^2(-9+4T\overline{p}(\rho-1)+7\rho)\kappa\end{pmatrix}}{((1+Ta)(1+2Ta)(Ta(-1+\rho)-1)\kappa)} \quad \text{(B3)}$$

$$\frac{\sigma^2_{O,gOUT}}{\sigma^2_{D_{ARMA}}}=\frac{-\begin{pmatrix}2Ti^3(\rho-1)\kappa-2T\overline{p}(1+T\overline{p})(\alpha-1)(\rho^2-1)+Ti^2(4T\overline{p}(\rho-1)-3+\rho)\kappa+\\ 2Ta^3(\rho-1)(2-2\rho+\alpha(Ti\alpha-2+(2+(Ti-1)\alpha)\rho))+\\ Ta^2(-3+6Ti(\rho-1)+4T\overline{p}(\rho-1)+\rho)(2-2\rho+\alpha(-2+Ti\alpha+(2+(Ti-1)\alpha)\rho))+\\ Ti\begin{pmatrix}2T\overline{p}(\alpha-2)^2(\rho-1)+2(2\alpha+\rho-2)+\\ \rho(\alpha^2+2\rho-2\alpha(1+\rho))+2T\overline{p}^2(\rho-1)\kappa\end{pmatrix}+\\ Ta\begin{pmatrix}4T\overline{p}(2+T\overline{p})(\alpha-1)+(\alpha-2)(2T\overline{p}(1+T\overline{p})(\alpha-2)+\alpha)\rho-\\ 2T\overline{p}(\alpha^2+T\overline{p}(2+(\alpha-2)\alpha))\rho^2+2(\alpha+\rho-1)+2Ti^3\alpha^2(\rho^2-1)+\\ Ti(\rho-1)\begin{pmatrix}2T\overline{p}^2\alpha^2(1+\rho)-2(\rho-7)+\alpha(\alpha-14+2\rho-3\alpha\rho)-\\ 2T\overline{p}(8(\rho-1)+\alpha(8-3\alpha+(\alpha-8)\rho))\end{pmatrix}+\\ Ti^2\left(12\alpha(\rho-1)^2-12(\rho-1)^2+\alpha^2\left(\rho(4+\rho)-9+4T\overline{p}(\rho^2-1)\right)\right)\end{pmatrix}\end{pmatrix}}{((1+2Ta)(Ta+Ti)(2Ti-1)(-1+Ta(\rho-1))(Ti(\rho-1)-\rho)\kappa)} \quad \text{(B4)}$$

B.4 Net stock variance ratio

Eq (B5) is the resulting closed form for the inventory variance produced by the generalised OUT policy in response to ARMA demand. Notice this closed form is valid for all lead-times (*Tp*) and inventory feed-forward (*a*) and feedback gains (*Ti*). This was arrived at by using the same procedure as the bullwhip expressions, except we started by manipulating Figure B3 for the Net Stock transfer function. The expression is also very lengthy, thus we have made the following substitutions in (B5); $T\overline{p}=Tp+a$, $\Xi=(1+Ta)$, $\Phi=(Ti-1)$, $\Psi=(\rho-1)$, $\Omega=(1+Ti\Psi)$, $\Theta=(\Xi-Ti)$, $\chi=(\rho+Ta\Psi)$. None of these substitutions contain *Tp*, but the following do; $\varphi=(\alpha+\Psi)\left((1-\rho)(T\overline{p}+Ta+Ti)+\rho^{Tp}\chi\Omega-\rho\right)Ta^{Tp}$, $\zeta=Ta^{-Tp}\rho\Phi^{-Tp}$, $\phi=(T\overline{p}+Ti)(\alpha+Ta\alpha-1)\Xi^{Tp}\rho^{Tp}\Omega$, $\tau=Ta^{Tp}Ti^{Tp}(Ti\alpha-1)(T\overline{p}+\Xi)\rho^{Tp}\chi$.

$$\frac{\sigma^2_{NS_{gOUT}}}{\sigma^2_{D_{ARMA}}} = \frac{\left(\left(\rho^2-1\right)\left(\frac{\left(Tp\alpha^2\left(\rho^2-1\right)-1+\rho\left(\begin{array}{c}2+2\alpha(\rho-1)-\rho+\alpha^2(2+\rho)-2\alpha\rho^{Tp}(1+\rho)\\(\alpha+\rho-1)+\rho^{1+2Tp}(\alpha+\rho-1)^2\end{array}\right)\right)}{(\rho-1)^3(1+\rho)}+\frac{\left(\zeta^2\left(\frac{1}{Ti^2\Xi^2}\right)^{Tp}\left(\begin{array}{c}\frac{\tau^2\Phi^2(\Xi\Phi)^{2Tp}}{Ti^2-\Phi^2}+2\tau\Phi^{1+Tp}(\Xi\Phi)^{Tp}\left(\frac{Ta(TaTi)^{Tp}\phi}{Ta\Phi-Ti\Xi}+\frac{\Theta\rho(Ti\Xi\rho)^{Tp}\varphi}{\rho\Phi\Psi-Ti\Psi}\right)+\Phi^{2Tp}\\\left(Ta(TaTi)^{Tp}\phi\left(\frac{Ta(TaTp)^{Tp}\phi}{\Xi^2-Ta^2}+\frac{2\Theta\rho(Ti\Xi\rho)^{Tp}\varphi}{\Xi\Psi-Ta\rho\Psi}\right)-\frac{\Theta^2\rho^2(Ti\Xi\rho)^{2Tp}\varphi^2}{(\rho^2-1)\Psi^2}\right)\end{array}\right)\right)}{\Theta^2\chi^2\Omega^2}\right)\right)}{-\left(2-2\rho+\alpha(\alpha+2\rho-2)\right)} \quad \text{(B5)}$$

All of the variance ratios in this paper have been incorporated into a Microsoft Excel Add-in that is available upon request.

Acknowledgements

This research was supported by the Fund for Scientific Research-Flanders (Belgium) under project FWO G.0051.03 and by the Cardiff Young Researchers Initiative. We would like to thank our referees for their insightful comments and suggestions.

References

Axsäter, S. 1985. Control theory concepts in production and inventory control. *International Journal of Systems Science,* **16** 2 161-169.

Baganha, M.P., Cohen, M.A. 1998. The stabilizing effect of inventory in supply chains. *Operations Research* **46** 3 572-583.

Balakrishnan, A., Geunes, J., Pangburn, M.S. 2004. Coordinating supply chains by controlling upstream variability propagation. *Manufacturing & Service Operations management* **6** 2 163-183.

Box, G.E. and Jenkins, G.M. 1970. Time series analysis forecasting and control. Holden-Day, San Francisco.

Brown R.G. 1962. Smoothing, forecasting and prediction of discrete time series. Prentice-Hall, Inc. Englewood Cliffs, N.J.

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

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R. 2003. Measuring the Bullwhip Effect: A control theoretic approach to analyse forecasting induced Bullwhip in order-up-to policies. *European Journal of Operations Research* **147** 3 567-590.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R. 2004. The impact of information enrichment on the Bullwhip effect in supply chains: A control engineering perspective. *European Journal of Operational Research,* **153** 727-750.

Deziel, D.P., Eilon, S. 1967. A linear production – inventory control rule. *Production Engineer* **43** 93-104.

Disney, S.M. and Towill, D.R. 2003. On the bullwhip and inventory variance produced by an ordering policy. *OMEGA: The International Journal of Management Science* **31** 3 157-167

Disney, S.M., Towill, D.R. 2002. A discrete linear control theory model to determine the dynamic stability of Vendor Managed Inventory supply chains. *International Journal of Production Research* **40** 179-204.

Forrester, J. 1958. Industrial Dynamics, A major breakthrough for decision makers. *Harvard Business Review*, July-August 67-96.

Graves, S.C. 1999. A single-Item inventory model for a non-stationary demand process. *Manufacturing and Service Operations Management* **1** 50-61.

Holmström, J. 1997. Product range management: a case study of supply chain operations in the European grocery industry. *Supply Chain Management* **2** 3 107-115.

John, S., Naim, M.M., Towill, D.R., 1994. Dynamic analysis of a WIP compensated decision support system. *International Journal of Manufacturing Systems Design* **1** 4 283-297.

Lee, H.L., Padmanabhan, P., Whang, S. 1997b. The bullwhip effect in supply chains. *Sloan Management Review* Spring 93-102.

Lee, H.L., Padmanabhan, P., Whang, S., 1997a. Information distortion in a supply chain: the Bullwhip Effect. *Management Science* **43** 543-558.

Lovejoy, W. 1992. Stopped myopic policies in some inventory models with generalized demand processes. *Management Science* **38** 5 688-707.

Magee, J.F. 1956. Guides to inventory policy-II. Problems of uncertainty. *Harvard Business Review,* March-April, 103-116.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R., & Winkler, R. 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**, 11-153.

Muth, J.F. 1960. Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association* **55** 299-306.

Nise, N.S. 1995. Control systems engineering. The Benjamin Cummings Publishing Company, Inc., California.

Rao, U. 2003. Properties of the periodic review (R, Q) inventory control policy for stationary stochastic demand. *Manufacturing & Service Operations Management* **5** 1 37-53.

Schmenner, R. W. 2001. Looking ahead by looking back: Swift, even flow in the history of manufacturing. *Production and Operations Management* **10** 1 87-95.

Silver, E.A., Pyke, D.F., Peterson, R. 1998. Inventory management and production planning and scheduling. John Wiley & Sons, New York.

Simon, H.A. 1952. On the application of servomechanism theory to the study of production control. *Econometrica* **20** 247-268.

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

Towill, D.R. and Yoon, S.S. 1982 Some features common to inventory system and process controller design. *Engineering Costs and Production Economics* **6** 225-236.

Treharne, T. and Sox, C. 2002 Adaptive inventory control for non-stationary demand and partial information. *Management Science* **48** 607-624

Tsypkin Y.Z. 1964. Sampling systems theory and its application. Pergamon Press, Oxford.

Vassian H.J. 1955. Application of discrete variable servo theory to inventory control. *Journal of the Operations Research Society of America* **3** 272-282.

Zipkin, P.H. 2000. Foundations of inventory management. McGraw-Hill, New York.

---

[1] We may use the terms production rates, order rates, purchase orders and replenishments interchangeably depending on context. For example if our OUT policy was placed in a factory context, production or order rates might be a descriptive term. Whereas for a distributor or retailer, purchase orders or replenishments might be more appropriate.

[2] Interestingly, Vassian, Magee and Brown all worked at the Boston based OR consulting company, Arthur D. Little.

[3] In this contribution, the closed form variance ratios are independent of the actual distribution of the white noise process driving the demand generator. However, for the fill-rate expressions used later in our story we have assumed they are normally distributed.

[4] Differentiating $Bullwhip + NSAmp = \frac{Ti^2 - Tp + 2TiTp + 1}{2Ti - 1}$ with respect to *Ti* yields $\frac{2Ti(Ti-1)-2}{(1-2Ti)^2}$, solving for zero gradient and selecting the relevant root yields the optimum *Ti* to minimise the sum of bullwhip and inventory variance. It is, $\frac{1+\sqrt{5}}{2}$, that we will recognise as the "Golden Ratio".

[5] Thus any analysis where the target inventory level is a function of demand will not be tractable as the inventory levels will also have infinite variance and no natural mean. Neither will an analysis of bullwhip. However, studies on the inventory variance in a constant target inventory system are possible for a non-stationary demand, for example see Graves (1999).

[6] Confirming managerial expectations, see Forrester (1958), and re-emphasizing the time-compression paradigm.