Dejonckheere, J., Disney, S.M., Lambrecht, M.R. and Towill, D.R., (2003), "Measuring and avoiding the bullwhip effect: A control theoretic approach", European Journal of Operational Research, Vol. 147, No. 3, pp567-590. DOI: 10.1016/S0377-2217(02)00369-7.

Measuring and avoiding the bullwhip effect: A control theoretic approach

J. Dejonckheere[1], S. M. Disney[2], M. R. Lambrecht[3] and D.R. Towill[2]

(1) GE Power Controls Belgium,
Nieuwevaart 51,
9000 Ghent, Belgium.
Jeroen.Dejonckheere@gepc.ge.com
Formally (3)

(2) Logistics Systems Dynamics Group,
Cardiff Business School,
Cardiff University,
Aberconway Building, Colum Drive,
Cardiff – UK, CF10 3EU. Wales, UK
DisneySM@Cardiff.ac.uk

(3) Department of Applied Economics,
Katholieke Universiteit Leuven,
Naamsestraat 69,
B – 3000 Leuven, Belgium.
Marc.Lambrecht@econ.kuleuven.ac.be

Abstract

An important contributory factor to the bullwhip effect (i.e. the variance amplification of order quantities observed in supply chains) is the replenishment rule used by supply chain members. First the bullwhip effect induced by the use of different forecasting methods in order-up-to replenishment policies is analysed. Variance amplification is quantified and we prove that the bullwhip effect is guaranteed in the order-up-to model irrespective of the forecasting method used. Thus, when production is inflexible and significant costs are incurred by frequently switching production quantities up and down, order-up-to policies may no longer be desirable or even achievable. In the second part of the paper a general decision rule is introduced that avoids variance amplification and succeeds in generating smooth ordering patterns, even when demand has to be forecasted. The methodology is based on control systems engineering and allows important insights to be gained about the dynamic behaviour of replenishment rules.

Keywords

Supply Chain Management, replenishment rule, bullwhip effect, production smoothing, system dynamics

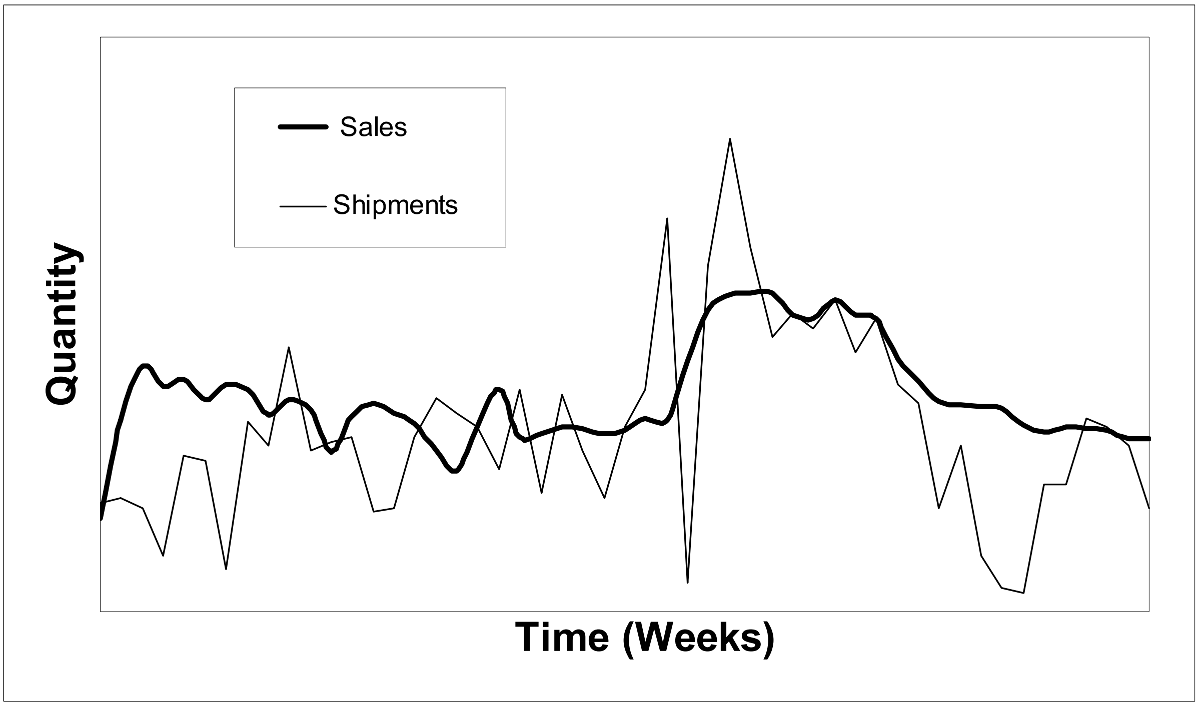
1. Introduction

The tendency of orders to increase in variability as one moves up a supply chain is commonly known as the bullwhip effect. Forrester (1958, 1961) initiated analysis of this variance amplification phenomenon. His work has inspired many authors to develop business games to demonstrate the bullwhip effect. The well-known Beer Game originated from MIT at the end of the fifties and Sterman (1989) reports on the major findings from a study of the performance of some 2000 participants. Kaminsky and Simchi-Levi (1998, 2000) developed a computerised version of the beer game. There is certainly no lack of empirical evidence from real world supply chains. The figures below illustrate the bullwhip effect observed in the real world. They show data from a major retailer and a manufacturer of fast moving consumer goods. In figure 1a we show the retailer sales versus the shipments from the manufacturers' distribution centre (to this retailer) for one specific product. The shipments are clearly much more variable than the sales. In figure 1b we show the shipments from the manufacturing plant to the distribution centres versus the production quantities. Again we observe a drastic increase in variability. Our data shows that while coefficients of variation[1] (CV's) of retail sales typically range between 0.15 and 0.5, the CV of the production orders are typically in the range of 2 to 3. In other words, a very substantial increase in variance has occurred.

Lee, Padmanabhan and Whang (1997a, b) identify five major causes of the bullwhip effect: the use of 'demand signal processing', nonzero lead times, order batching, supply shortages and price fluctuations. In this paper, we will mainly focus on the issue of non-zero lead-times and particularly demand signal processing. We understand demand signal processing as the practice of adjusting the demand forecasts and as a result of this practice, adjusting the parameters of the inventory replenishment rule. Doing this may cause over-reactions to short-term fluctuations and lead to variance amplification. Baganha and Cohen (1998) correctly formulated a very puzzling idea. We quote: "inventory management policies can have a destabilising effect by increasing the volatility of demand as it passes up to the chain" whereas "one of the principal reasons used to justify investment in inventories is its role as a buffer to absorb demand variability". In other words, inventories should have a stabilising effect on material flow patterns. How is it that market variability is amplified rather than dampened? We show that the design of inventory replenishment rules plays a crucial role in that respect. Given the common practice in retailing to replenish inventories very frequently (e.g. daily) and the tendency of manufacturers to produce to demand, we will focus our analysis on a class of replenishment strategies known as order-up-to level policies. In the absence of fixed ordering costs, it is known that it is optimal to bring the inventory position up to a predetermined target level. This is the simplest form an optimal ordering rule can take.

An order-up-to policy is optimal in the sense that it minimises the expected holding and shortage costs. We refer to the important work of Samuel Karlin (1958) for the theoretical foundation of this approach. However, bullwhip effect research is also interested in the control of inventory and (production or distribution) order rate fluctuations. In this paper we therefore analyse the behaviour of order-up-to policies in terms of order-rate fluctuations. In

[1] The Coefficient of variation (CV) of a dataset is the ratio of the standard deviation of the data over the mean.



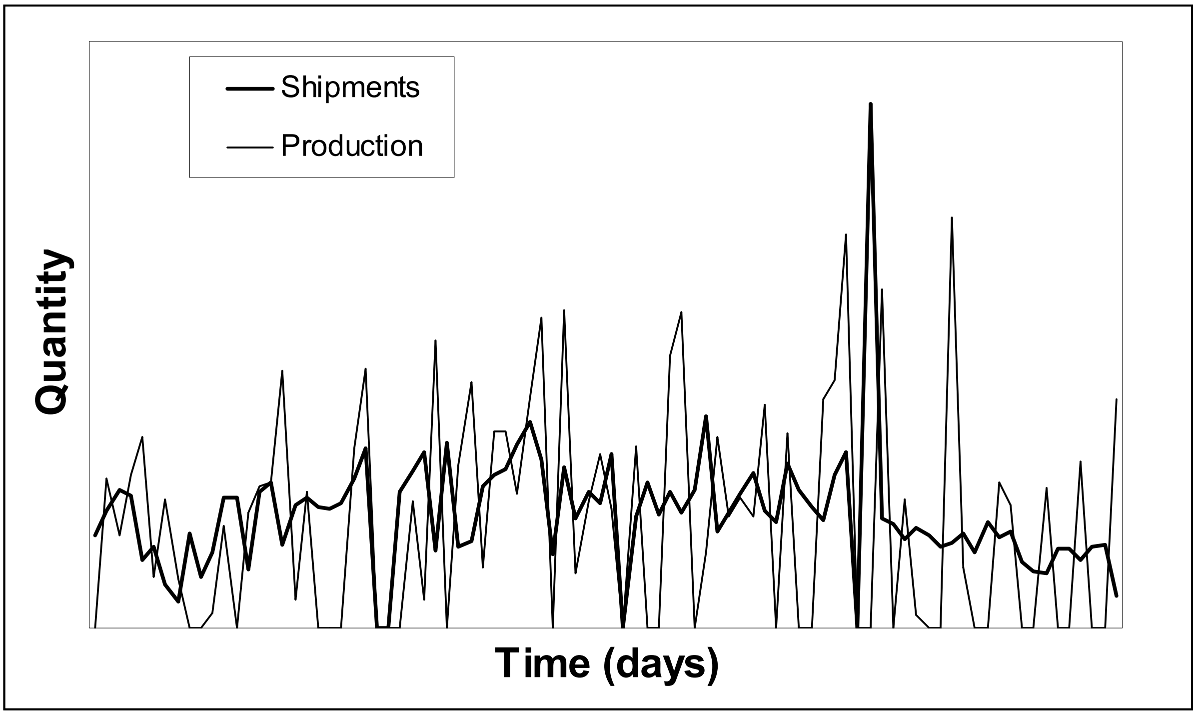


Figure 1a and 1b: The Bullwhip Effect: real-life illustration[2]

practice, production may be very inflexible and significant costs may be incurred by ramping up and down production levels frequently. In the case where we also penalise fluctuations in

[2] Note. Figure 1a shows weekly data between the manufacturer and one particular retailer. This plot is appropriate as the retailer places orders weekly. However Figure 1b shows the daily production data for the same product, but for a number of retailers (of which one is the retailer in Figure 1a). This plot is appropriate as the production planning is done on a daily basis.

production or ordering quantity levels, order-up-to policies may cease to be optimal. These production-switching costs may be very large, consequently requiring a smooth production or ordering pattern. In the second part of the paper, we present a general decision rule that can be used to generate smooth ordering patterns. Hence, although the inventory related costs can be increased by smoothing the demand pattern, the decision rule may still outperform order-up-to policies in terms of total costs (inventory holding and shortage costs plus production switching costs). This will of course depend on the particular cost structure of the supply chain under consideration.

We distinguish between two methodological approaches to tackle the problem. One is the statistical inventory control approach, the other one being the control systems engineering approach. In this paper we advocate the engineering approach and are able to confirm and extend the existing results obtained through statistical analysis. We now briefly review the recent major contributions of both methodologies. Lee, Padmanabhan and Whang (1997a, b) and Chen, Drezner, Ryan and Simchi Levi (2000a, b) use statistical approach to quantify the bullwhip effect. They quantify the impact of exponential smoothing based forecasts on order-rate fluctuations within order-up-to policies. Lower bounds are established for the variance amplification in a simple supply chain consisting of a single manufacturer and a single retailer. Several types of demand processes are assumed. The authors are able to provide important managerial insights, such as the fact that the bullwhip effect is caused by the need to forecast and the smoother the demand forecasts, the smaller the bullwhip effect.

Application of control engineering to production and inventory control was first achieved by Simon (1952) by using the Laplace transform. This move was quickly translated into the newly favoured discrete z-domain by the OR community, mostly notably by Vassian (1955), Adelson (1966), Elmaghraby (1966) and Deziel and Eilon (1967). It is noticeable from a literature search that contributions that utilise the Laplace transform are more numerous then those utilising the z-transform. This is probably due to the more tractable algebraic manipulation required when using the Laplace transform. Unfortunately, the order-up-to model is inherently discrete, forbidding the use of the Laplace transform. However, as the z-transform is a special case of the Laplace transform, many tools, techniques and best practises developed for the Laplace transform are readily exploited in the z-domain – usually after a small change in notation. We also refer to Towill (1970, 1982), Bertrand (1986), Bonney and Popplewell (1988), Bonney, Popplewell and Matoug (1994), Grubbström (1998), Towill (1999), Dejonckheere, Disney, Lambrecht and Towill (2001) and Disney (2001) for contributions on replenishment rules and inventory fluctuations using transform techniques. The works of Aseltine (1958), Jury (1964) and Houpis and Lamont (1985) are particularly illuminating texts describing the mathematics of the z-transform.

A brief literature review shows that replenishment rules have largely been analysed by the;

- OR community by exploiting the z-transform directly,
- System dynamics community by generating simulations from causal loop diagrams, although knowledge of control theory is often advocated as a useful source of inspiration,
- Control theory community by using signal flow diagrams, block diagrams, s/z-transforms, "hard system" control laws, frequency response plots and simulation.

Thus, the presentation of this research via causal loop diagrams, block diagrams, z-transforms, frequency response plots and simulation will be directly accessible and relevant to a large audience. We refer to Sterman (2000) for an overview of the systems thinking approach.

In this paper we will measure the variance amplification of orders within order-up-to policies from a control engineering perspective. We will prove that classical order-up-to policies will always generate a bullwhip effect. It is however possible to design replenishment rules based on what we call fractional adjustments, thereby generating smooth order patterns. In other words, it is possible to dampen order fluctuations even in environments where decision makers have to rely on forecasts.

The remainder of the paper is organised as follows: in the second section, an overview of the control engineering based methodology is given. In the third section, we analyse the bullwhip effect induced by exponential smoothing forecasts in order-up-to policies. In the fourth section, we measure variance amplification for two other types of forecasting methods within the order-up-to setting: moving averages and demand signal processing (Lee et al., 1997b and 2000) and we compare those forecasting systems with the exponential smoothing scenario of section 3. In the fifth section, a general replenishment rule is proposed that can be used to generate smooth ordering patterns for different demand patterns. Finally, a summary is given of all the policies analysed and insights presented in this paper.

2. Methodology

The methodology used in this paper is control systems engineering (transfer functions, frequency response curves and spectral analysis). We complement this methodology with simple spreadsheet analysis in order to test our insights via system responses to real-life demand patterns and to perform the Fast Fourier Transform on real-life demand patterns. First we briefly review the most important concepts and techniques.

a. Transfer function

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 to its input and in this paper is defined as the ratio of the z-transform of the output variable to the z-transform of the input variable. Since supply chains can be seen as systems, with complex interactions between different parts of the chain, we can use a transfer function approach to model these interactions. For every replenishment rule, a transfer function will be developed that completely represents the dynamics of this particular rule. Input to the system corresponds to the demand pattern and output refers to the corresponding replenishment or production orders. For more details on control engineering and transfer functions, we refer the reader to the appropriate literature, although Nise (1995) provides a good introduction. In sections 3 and 4, we will illustrate how we obtain the transfer function by constructing the 'causal loop diagram' and the 'block diagram' for the replenishment rule under consideration.

b. The frequency response plot

To derive the 'frequency response' plot (FR) of a replenishment rule, we will present the rule with sinusoidal inputs of different frequencies; that means we want to know what orders (output) are generated when the demand (input) is sinusoidal. Since we are dealing with linear systems, we know that in the steady state the output will also be a sine wave with the same frequency, but the amplitude and the phase angle may have changed. We will be particularly interested in the ratio of the amplitude of the generated orders (output) over the amplitude of the sinusoidal demand (input): this is known as the Amplitude Ratio (AR). We will present the replenishment rule with sine waves of frequencies ranging from 0 to π radians per sample interval. For all of these frequencies, we can find the AR and in this way draw the FR plot and hence have an extremely insightful profile of the system dynamics. Because of the fact that real life demand data can be seen as composed of different sinusoids, it is intuitive to analyse responses to different sine waves. The FR will be used to make predictions on whether or not, and to what extent, the replenishment rule will lead to variance amplification. Consequently, some new metrics for the bullwhip effect are introduced specifically based on the FR plot.

c. Spectral analysis

Spectral analysis is a mathematical technique used to decompose a time series into constituent frequencies or periodicities. The amplitude or variance associated with each frequency component is known as the ‘spectral density estimate.’ The Fourier transform, Cochran et al (1967), is an algebraic method of decomposing any time series into a set of pure sine waves of different frequencies, with a particular amplitude and phase angle associated with each frequency. The algebraic sum of the sinusoidal components, adjusted for phase angle will accurately reproduce the original time series. There are a variety of methods to calculate the spectral density estimates. A technique often used is called the Fast Fourier Transform (FFT), Cochran et al (1967). This technique greatly reduces the time required to perform a Fourier analysis on a computer, and can be obtained by using simple spreadsheets. We will illustrate the FFT on the demand pattern shown by the shipment data in figure 1b and this is designated ‘demand’ in this paper.

We have a demand history of 128 periods: this input can be seen as the sum of a constant term and plus 63 sine waves of increasing frequency (in integer multiples of the base frequency (1/128)2π radians per sample interval). If all the individual sine waves were added to the constant term, we can exactly reconstruct the original data set (see Makridakis, 1978). The amplitudes of the 63 sine waves are a measure of their relative importance in recomposing the original data set. The amplitudes are given in figure 2. The plot of the amplitude versus the frequencies is called the ‘periodogram’ of the demand pattern under consideration.

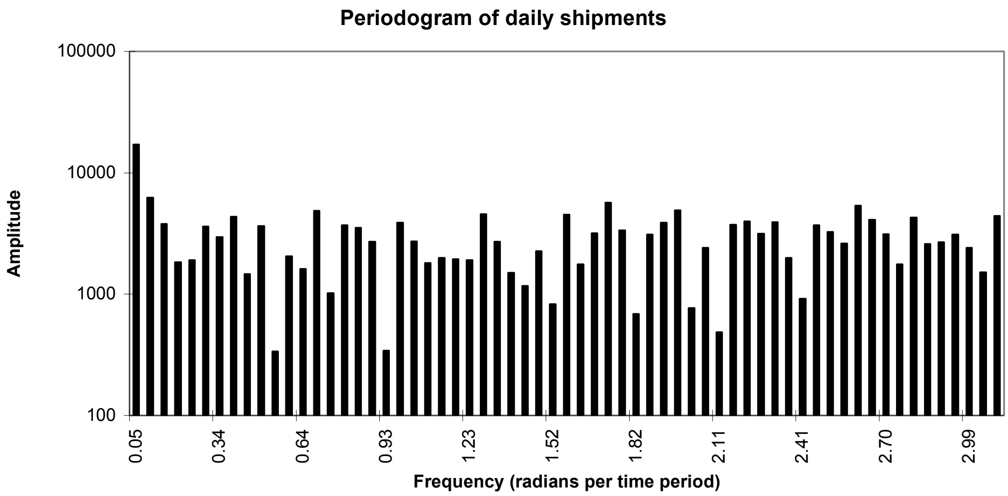


Figure 2: Amplitudes of the different sinusoids in the demand pattern given in the shipments in figure 1b

The control engineering based methodology used in this paper is summarised in the figure 3 below. Replenishment rules used in supply chains can be described by means of transfer functions. Based on the transfer function, the FR plot can be obtained. Next, spectral analysis will be applied to demand patterns resulting in a periodogram. This information can be used to measure and predict the magnitude of the bullwhip effect.

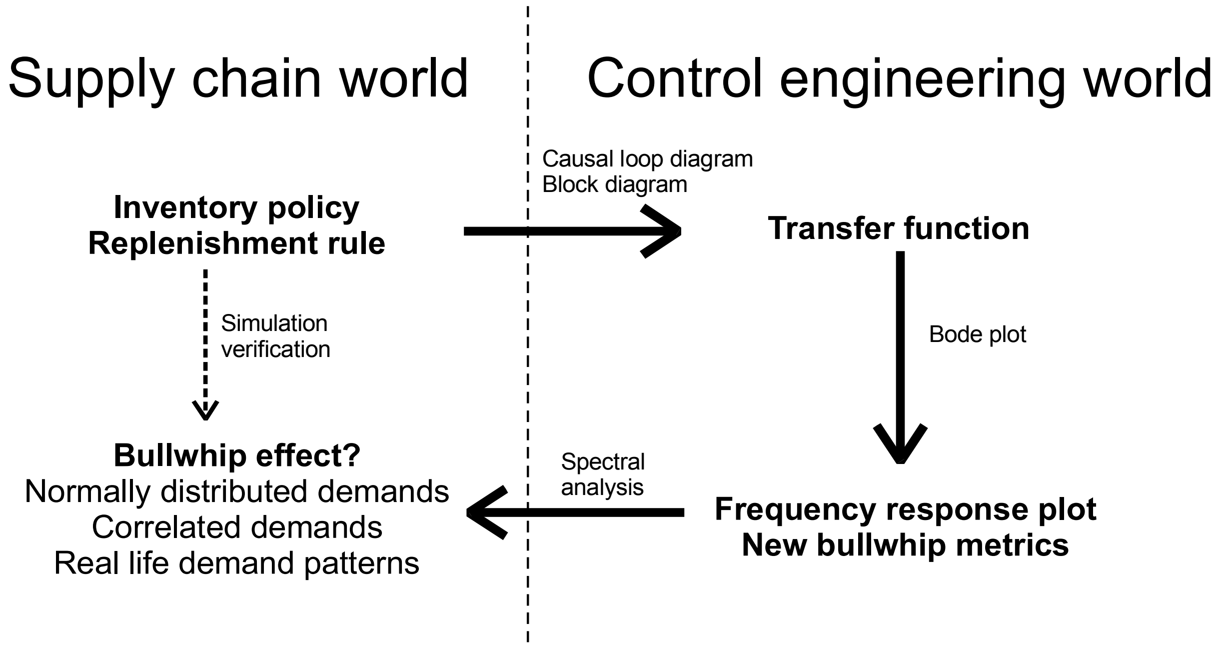


Figure 3: Summary of control engineering based methodology

3. The bullwhip effect caused by order-up-to policies based on exponential smoothing forecasts

Consider a simple supply chain consisting of a single retailer and a single manufacturer. We assume the following sequence of events: in each period t, the retailer first receives goods, then demand D_t is observed and satisfied (if not backlogged), next, the retailer observes the new inventory level and finally places an order O_t on the manufacturer. Any unfilled demand is backlogged in our model. There is a fixed lead time between the time an order is placed by the retailer and when it is received at the retailer, such that an order placed at the end of period t is received at the start of period $t+L$. Specifically, the lead-time L, consists of one time period ordering delay and T_p time periods of physical production or distribution delay. Thus when L=1 the production/distribution lead-time = 0, i.e. the production/distribution is instantaneous, i.e. if a product with a production/distribution lead-time of zero is ordered in time period t, it arrives in time period t+1. This is because, of course, the receipt of the order is delayed by one period because of the sequence of events. We assume in our numerical examples used throughout the text that T_p equals 3 time periods. There is nothing special in this choice of delay.

In this section, we first analytically describe the order-up-to policy based on exponential smoothing forecasts. Then we derive the transfer function and determine the FR. Furthermore, with the FR plot, we will highlight some interesting insights into the bullwhip effect created by using this replenishment rule. And finally, we will look at the impact of the smoothing parameter α on the resulting variance amplification.

a. The decision rule

In any order-up-to policy, ordering decisions are as follows:

$$O_t = S_t - \textit{inventory position}_t \tag{1}$$

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 $S_t = \hat{D}_t^{\ L} + k\ \hat{\sigma}_t^{\ L}$ (2)

where $\hat{D}_t^L$ is an estimate of mean demand over *L* periods ($\hat{D}_t^L = L\hat{D}_t$), $\hat{\sigma}_t^L$ is an estimation of the standard deviation of the demand over *L* periods, and *k* is a chosen constant to meet a desired service level.

Chen et al. (2000a) correctly mention that when the average and the standard deviation of the demand during the lead-time are *known* with certainty, the order-up-to level would be constant, and in every period, the retailer would order the last observed demand. Hence, there is no bullwhip effect. However, when D^L and σ^L are unknown, the retailer must forecast demand. This forecasting creates variability in the order-up-to level and causes the bullwhip effect. Thus, for every period, the retailer updates the order-up-to level with the current estimates. To simplify the analysis (and to ensure that our model matches the model of Chen et al (2000a) – to be illuminated later at a more appropriate point in the discussion), we have set *k* equal to zero and increased the lead-time by one. Policies of this form are often used in practice: the value of *L* is inflated and the extra inventory represents the safety stock. In other words *L* not only represents the physical lead-time, but also a safety lead-time. Remember that *L* already includes a nominal one period order delay because of the sequence of events, so that we now have $L = T_p + 2$.

In this section, we use simple exponential smoothing to forecast demand. The formula for simple exponential smoothing is well known to be:

$$\hat{D}_t = \hat{D}_{t-1} + \alpha * (D_t - \hat{D}_{t-1}) \text{ or } \hat{D}_t = \alpha * D_t + (1-\alpha) * \hat{D}_{t-1} \tag{3}$$

Note that since we make the ordering decision at the end of the period, the current demand D_t can be used in the forecast $\hat{D}_t$. For simple exponential smoothing, the average age of the data in the forecast is equal to $(1-\alpha)/\alpha$ (Makridakis, 1978). Let T_a be the average age of the data in the forecast, consequently $\alpha = 1/(1+T_a)$.

b. Deriving the transfer function

In order to derive the transfer function for a particular order-up-to policy, we first have to draw the 'causal loop diagram'. This is shown in figure 4. We refer to Sterman (2000) for a useful tutorial on constructing and interpreting causal loop diagrams. Causal loop diagrams are a helpful means of communication, but need translating into a rigorous block diagram, which is our next step.

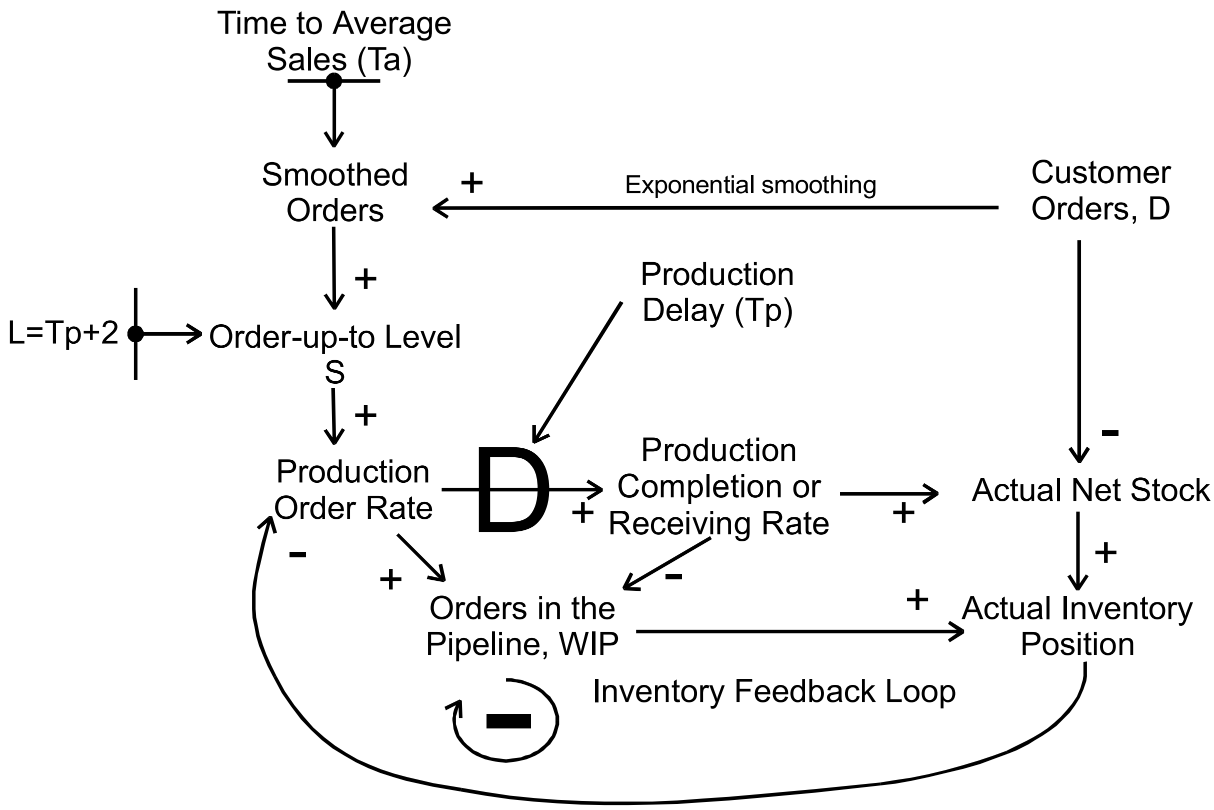


Figure 4: Causal loop diagram for the order-up-to policy based on exponential smoothing forecasts

The corresponding block diagram is given in the figure 5. Note that there is a delay operator in the block diagram to ensure the correct sequence of events as mentioned earlier. If, this delay were omitted in a spreadsheet simulation, the software would return an error message because of the existence of an algebraic loop, which is obviously not a physically realisable situation[3]. Applying the well-established rules for block diagram reduction, we obtain the following transfer function for the orders made under policy defined by equations (1-3) *O*, over the observed demand *D*:

$$\frac{O}{D} = \frac{-(2 + T_p + T_a) + (3 + T_p + T_a)z}{T_a(-1 + z) + z} \tag{4}$$

To derive this transfer function in equation (4), we had to use the *z*- transform for the exponential smoothing algorithm $F(z) = \frac{\hat{D}}{D} = \frac{\alpha}{1-(1-\alpha)z^{-1}}$, (Wikner, 1994). (5)

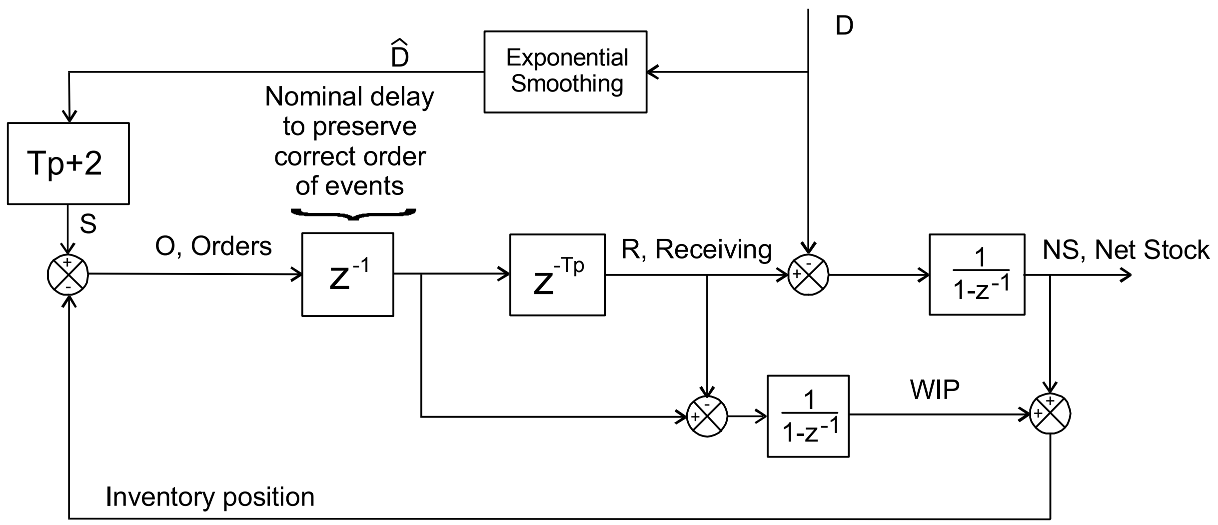


Figure 5: Block diagram for the order-up-to policy with exponential smoothing forecasts

Remember that our sequence of events was the following: 1. receive, 2. satisfy demand, 3. count inventory, 4. place order. It is important that in this sequence of events the order is made after the demand has been realised and fulfilled. This is also the case for Lee, So, Tang (2000), where their ordering decision at the end of the ordering period equals;

$O_t = D_t + S_t - S_{t-1}$ (6)

[3] Additionally, if demand is known with certainty then the order-up-to model is expected to produce orders that are the same as the last observed demand, i.e. to simply pass on orders. If demand is known with certainty then Ta is set to ∞ (or a very large number) in the model to reflect that we don't have to update the estimated average demand since it is already known. If the nominal delay is omitted from this model then it produces an erroneous response. The pure order-up-to model should also have a perfectly flat frequency response of unity for all frequencies as it just passes on demand. If this delay is omitted from the transfer function then this property is also not obtained.

In such a scenario, in any time period t, the current demand D_t can already be used to update the forecasts and to calculate the order O_t made at the end of the time period. Furthermore, if the order-up-to is constant, then $O_t=D_t$.

However, another type of sequence of events is also used in the literature and the crucial difference is that the order is placed before the demand has been observed, e.g. at the beginning of the time period, (see Lee, Padmanabhan, Whang (1997b) and Chen, Ryan and Simchi-Levi (2000a)). In these cases, the order decision O_t is based on the previous periods demand and inventory position:

$$O_t=D_{t-1}+S_t-S_{t-1} \tag{7}$$

For a constant order-up-to level, we then have $O_t=D_{t-1}$. In general we may say that for a constant order-up-to levels, in both scenarios the order equals the last observed demand. So although there are two different sequences of events and there are two different forms of notation in the literature, both systems are actually the same as their differences “cancel out”. This has been exploited by Chen, Drezner, Ryan and Simchi-Levi (2000), who have also correctly used both sequences of events within their analysis.

c. The frequency response plot and new bullwhip metrics

For the order-up-to policy described by equations (1-3), we can now easily draw the FR plot. Technically, this is done by letting $z = e^{iw}$ in the transfer function of equation (4) and determining the size of the radius vector in the complex plane (i.e. its modulus). A discrete version of the FR can also be drawn in a spreadsheet by calculating AR-values for a series of sine waves with the frequencies gradually increasing from 0 to π radians per sample interval. The FR is given in figure 6, for the order-up-to policy defined by equations (1-3) with values of $T_a=8$ (or $\alpha=0.1111$), and with $T_p = 3$.

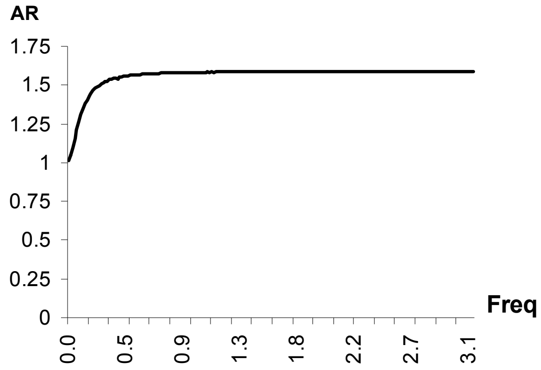


Figure 6: The FR for the order-up-to policy with exponential smoothing forecasts with Ta=8 and Tp=3

The conventional way to measure bullwhip is to use the ratio of the variance (or standard deviation or CV) of the orders being generated (output) to the variance (or standard deviation or CV) of the demand (input). In our control engineering methodology, we can define similar bullwhip metrics for any replenishment rule, based on its FR.

We could take the maximum of the FR plot as a bullwhip effect measure. For the order-up-to policy of equations (1-3), with T_p = 3 and T_a = 8, we have a peak AR–value of 1.588. This peak AR-value is a worse-case scenario, because it only occurs for a perfectly sinusoidal demand of one particular frequency. It can easily be shown that any *AR-value, is exactly the same as the ratio of the standard deviations of input over output*. Therefore, we need to know the arithmetic relationship between amplitude and variance of a sine wave: the square of the amplitude divided by two is equal to the variance (var = $A^2/2$, see Porges and Bohrer, 1991). The remainder of the proof is as follows:

$$AR = \frac{A_{output}}{A_{input}} = \frac{\sqrt{2 * \text{var (output)}}}{\sqrt{2 * \text{var (input)}}} = \frac{stdev\ (output)}{stdev\ (input)} \tag{8}$$

Since real demands are seldom perfectly sinusoidal, but rather a combination of different sine waves (see spectral analysis), we will use the area under the squared FR curve as another metric for the bullwhip effect. This second measure is a common metric in communications engineering (Garnell and East, 1977) and is called the "Noise Bandwidth". It is formally defined as $W_N = \int_0^{\pi} \left| F(e^{i\omega}) \right|^2 d\omega$, (9)

where $F(e^{i\omega})$ is the steady state response at excitation frequency ω. In particular cases the noise bandwidth may be calculated directly from formulae (Garnell and East, 1977).

Alternatively, if the FR plot is determined at discrete points, for example as in a spreadsheet, Noise Bandwidth can be estimated through numerical integration up to the Nyquist frequency (π radians per sampling period). For the order-up-to policy with smoothing parameter $T_a = 8$ and $T_p = 3$, we obtain $W_N = 7.625$.

d. Insights for supply chains

Based on both the FR plot and spectral analysis, we obtain four interesting insights concerning the bullwhip effect generated by replenishment rules in general, and by this order-up-to policy (1-3) in particular.

<u>Insight 1</u>: *The FR plot provides a valuable insight into the dynamic behaviour of the replenishment rule.*

Since this order-up-to policy overshoots for all possible frequencies from 0 to π radians per sampling period (see figure 6), and since every possible demand pattern is eventually a combination of different frequencies (spectral analysis), we know that this order-up-to policy will create variance amplification for every possible demand pattern. This first insight is very strong, because we do not have to make any assumptions regarding the distribution of demand!

<u>Insight 2</u>: *It is possible to quantify very accurately the magnitude of the variance amplification that will be caused by any replenishment rule when applied to any demand pattern, based on both the FR curve for the replenishment rule under consideration and the periodogram of this particular demand pattern.*

<u>Proof.</u> Suppose our demand pattern (input) is a time series I with N observations and suppose that N is an even number and a power of two. Applying the FFT, we decompose our input as the sum of a constant term C and (N/2-1) sine waves with frequencies that start at 0 radians per sampling interval and increase in integer multiples of a base frequency $f_0 = 1/N$ (Makridakis, 1978). The sine waves are denoted X_i with $i = 1, \ldots, (N/2)-1$. Adding up the sine waves results in the original input:

$$I = C + X_1 + X_2 + \ldots + X_{(N/2 - 1)} \tag{10}$$

Let the amplitude of the i^{th} sine wave be denoted A_i. Because the variance of the i^{th} sine wave is equal to $A_i^2/2$ and the covariance between two sine waves is always zero if the frequencies are different (Porges and Bohrer, 1991), the variance of our input signal can be written as:

$$\operatorname{var} I = 0.5\ A_1^2 + 0.5\,A_2^2 + ... + 0.5\,A_{(N/2-1)}^2 \ = 0.5 \sum_{i=1}^{N/2-1} A_i^2 \qquad \text{with } i = 1,\ldots,(N/2-1) \tag{11}$$

The physical meaning of equation (11) is that the spectral density estimate A_i^2 (also called the 'power') represents the contribution of frequency i to the total variance of the input signal (see also Chatfield, 1996). The generated orders (output) O based on policy (1-3) is again a summation of a constant term and (N/2 –1) sine waves, $\tilde{X}_i$,

$$O = C + \tilde{X}_1 + \tilde{X}_2 + ... + \tilde{X}_{(N/2\,-\,1)} \tag{12}$$

Since we are dealing with linear systems, the sine waves $\tilde{X}_i$ will have the same frequency as the sine waves X_i, but the amplitude and phase angle may have changed. Let $\tilde{A}_i$ determine the amplitude of sine wave $\tilde{X}_i$. To find the new amplitudes, we have to use the FR plot, which gives us AR-values for all frequencies from 0 to π radians per sampling interval. We will indicate the AR values corresponding with the (N/2 – 1) frequencies used in the spectral analysis as AR_i. Then we have $\tilde{A}_i = AR_i * A_i$, for i = 1, …, (N/2 – 1). Next, we can determine the variance of the generated orders as:

$$\operatorname{var} O = 0.5\ A_1^2 AR_1^2 + 0.5\,A_2^2 AR_2^2 + ... + 0.5\,A_{(N/2-1)}^2 AR_{(N/2-1)}^2 \ = 0.5 \sum_{i=1}^{N/2-1} A_i^2 AR_i^2 \tag{13}$$

Finally, the estimated variance amplification is given by the ratio of equation (13) divided by equation (11):

$$\frac{\operatorname{var} O}{\operatorname{var} I} = \frac{\sum_{i=1}^{N/2-1} A_i^2 AR_i^2}{\sum_{i=1}^{N/2-1} A_i^2}, \text{ with } i = 1, \ldots, (N/2-1) \tag{14}$$

(14) can also be rewritten as:

$$\frac{\operatorname{var} O}{\operatorname{var} I} = \sum_{i=1}^{N/2-1} \left(\frac{A_i^2}{\sum_{i=1}^{N/2-1} A_i^2} \right) * AR_i^2 \text{ ; with } i = 1, \ldots, (N/2-1) \tag{15}$$

Predicting the variance increase for any possible demand set is thus actually making a weighted average of all AR^2 values, with the weights being determined by the squared amplitudes found in the periodogram from the FFT.

Applying equation (15) to our illustrative demand pattern (the shipments in figure 1b) and using the amplitude values of figure 2 together with the amplitude ratios of figure 6, we obtain an estimated variance ratio of 2.036. To verify this result, we have developed a spreadsheet application enabling us to simulate various replenishment rules, allowing us to measure the variance amplification for any demand pattern (Lambrecht and Dejonckheere, 1999). We tested the previously described policy of equations (1-3) on the same demand pattern and we obtained a variance ratio of 2.047, illustrating the accuracy of our prediction. The generated orders according to equations (1-3) are given in figure 7 (dotted line):

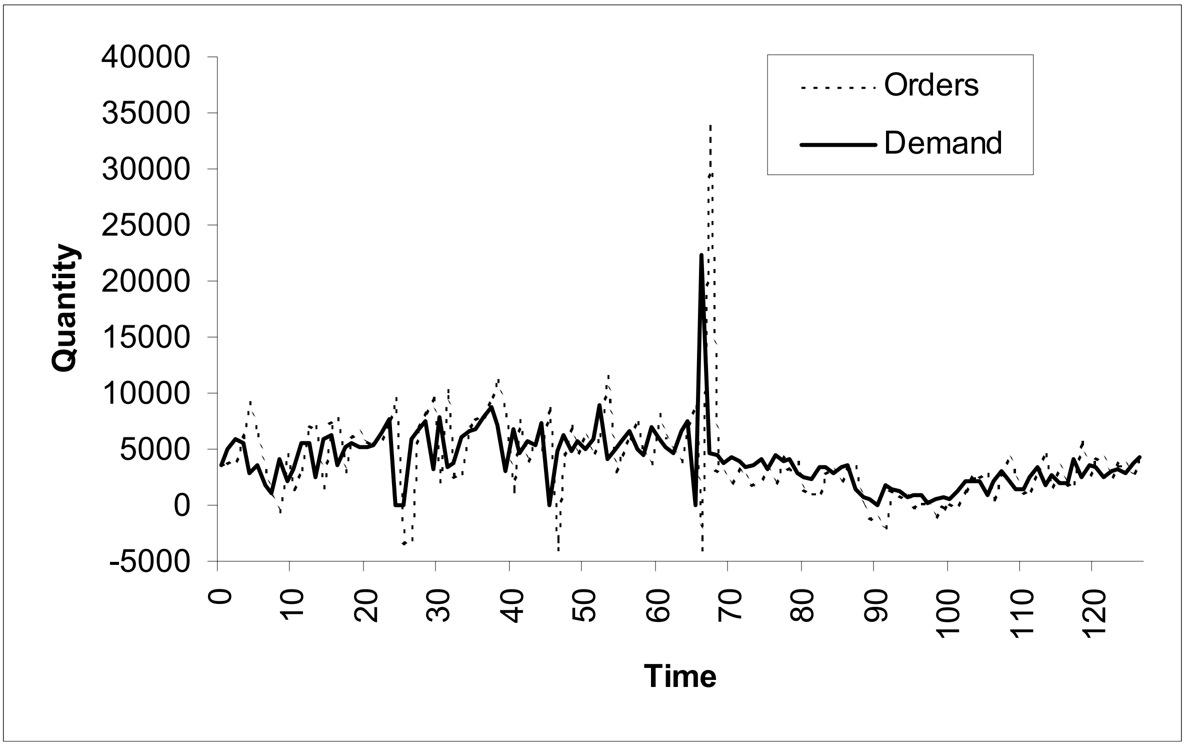


Figure 7: Simulated orders generated by the order-up-to policy defined equations (1-3) to the real-life shipments of figure 1b

In order to compare the estimated variance amplification using equation (15) with the simulated results, we selected 30 real life demand patterns (of 128 periods) from a manufacturer of fast moving consumer goods (data set 1 corresponds to the example used in this text and shown by the shipments of figure 1b). Column one of Table I summarises the results. Note that the average deviation is 0.279%.

	Col 1			Col 2			Col 3			Col 4		
Data sets	**Order up to policy based on exp. smoothing forecasts with T_a = 8**			**Order up to policy based on moving average forecasts with T_m = 17**			**Order up to policy based on demand signal processing with γ = 1**			**Fractional adjustment Smoothing decision rule with T_a = 8, T_n = T_w = 4**		
	predicted	**simulated**	**% gap**	**predicted**	**simulated**	**% gap**	**predicted**	**simulated**	**% gap**	**predicted**	**simulated**	**% gap**
Data set 1	2.0363	2.0445	0.4011	1.5790	1.6089	1.8584	3.7706	3.8754	2.7042	0.7534	0.7382	2.0646
Data set 2	2.4778	2.4759	0.0767	1.8552	1.8621	0.3716	4.4564	4.5342	1.7154	0.3749	0.3630	3.2782
Data set 3	2.1456	2.1433	0.1068	1.9106	1.9483	1.9340	1.8078	1.8586	2.7347	1.5013	1.4842	1.1549
Data set 4	2.4582	2.4598	0.0642	1.8240	1.7875	2.0421	4.9242	4.9566	0.6543	0.4401	0.4465	1.4290
Data set 5	2.3908	2.3887	0.0871	1.9065	1.8930	0.7116	4.3340	4.4527	2.6651	0.6820	0.6734	1.2846
Data set 6	2.3294	2.3323	0.1248	1.6925	1.7052	0.7455	4.8199	4.9353	2.3385	0.4826	0.4811	0.3072
Data set 7	2.3336	2.3391	0.2358	1.7428	1.7369	0.3423	4.3745	4.4181	0.9862	0.5925	0.5933	0.1389
Data set 8	2.4949	2.4973	0.0967	1.8311	1.8554	1.3075	5.1787	5.2246	0.8794	0.3096	0.3113	0.5498
Data set 9	2.4349	2.4401	0.2123	1.8014	1.7794	1.2359	5.0048	5.2102	3.9423	0.3908	0.3620	7.9654
Data set 10	2.2607	2.2475	0.5871	1.7133	1.7042	0.5367	4.5575	4.6169	1.2876	0.5612	0.4985	12.5866
Data set 11	2.4663	2.4671	0.0310	1.8067	1.8127	0.3326	4.4811	4.5232	0.9314	0.3680	0.3715	0.9541
Data set 12	2.4299	2.4268	0.1258	1.8045	1.7814	1.2957	4.7409	4.7581	0.3612	0.4401	0.4469	1.5225
Data set 13	2.4185	2.4264	0.3268	1.7710	1.7801	0.5106	4.2166	4.2791	1.4610	0.4669	0.4733	1.3501
Data set 14	2.4631	2.4624	0.0268	1.6885	1.7029	0.8476	4.8955	4.9766	1.6303	0.3331	0.3342	0.3229
Data set 15	2.1037	2.1220	0.8641	1.6553	1.6749	1.1706	3.9910	4.0548	1.5744	0.7533	0.7904	4.6914
Data set 16	2.2779	2.2782	0.0132	1.8231	1.8187	0.2408	3.6852	3.7263	1.1032	0.8540	0.8503	0.4304
Data set 17	2.4488	2.4501	0.0549	1.9371	1.9505	0.6890	5.1307	5.1476	0.3276	0.3166	0.3071	3.1102
Data set 18	2.4937	2.4659	1.1274	1.7520	1.7844	1.8166	5.4396	5.3168	2.3087	0.2689	0.3192	15.7687
Data set 19	2.4422	2.4444	0.0884	1.8429	1.8385	0.2401	4.8685	4.9023	0.6889	0.3448	0.3457	0.2617
Data set 20	2.4344	2.4262	0.3367	1.6509	1.6666	0.9414	5.8597	5.7712	1.5335	0.2768	0.2888	4.1544
Data set 21	2.2296	2.2282	0.0633	1.8475	1.8298	0.9651	3.1495	3.1818	1.0145	1.0145	1.0010	1.3456
Data set 22	2.4486	2.4503	0.0691	1.7836	1.8150	1.7299	5.7015	5.7871	1.4791	0.2811	0.2779	1.1697
Data set 23	2.4416	2.4343	0.2991	1.8359	1.8104	1.4070	5.7019	5.7814	1.3743	0.2733	0.2679	2.0046
Data set 24	2.4392	2.4388	0.0181	1.8426	1.8562	0.7327	6.0194	6.1347	1.8795	0.2377	0.2244	5.9275
Data set 25	2.4452	2.4470	0.0749	1.7797	1.8353	3.0268	5.6390	5.7404	1.7661	0.2863	0.2797	2.3700
Data set 26	2.4216	2.4287	0.2918	1.7777	1.7764	0.0725	4.2501	4.2959	1.0672	0.4656	0.4737	1.7095
Data set 27	2.4871	2.4885	0.0555	1.8146	1.7998	0.8251	4.2629	4.3654	2.3476	0.3840	0.3836	0.1121
Data set 28	2.4562	2.4633	0.2897	1.8018	1.8206	1.0352	4.3176	4.3491	0.7245	0.4490	0.4237	5.9647
Data set 29	2.2992	2.2574	1.8511	1.9900	1.9287	3.1777	3.1244	3.1157	0.2777	1.0248	0.9883	3.6896
Data set 30	2.2445	2.2358	0.3914	1.7500	1.8095	3.2895	4.1140	4.1568	1.0294	0.6765	0.6671	1.4121
average gap			**0.2797%**			**1.1811%**			**1.4929%**			**2.9677%**

Table I: Variance amplification for 30 real-life demand pattern data sets

Table I verifies that we can calculate the estimation of bullwhip (the coefficient of variation) from the frequency response via the frequency plot of the order-up-to system and the periodogram of the demand signal.

<u>Insight 3</u>: *Given the frequency response plot of a replenishment rule, the corresponding Noise Bandwidth divided by π (W_N / π) is equal to the expected variance increase (or decrease) for the replenishment rule applied to independently and identically distributed (i.i.d) normally distributed demands.*

The proof of insight 3 is relatively easy and intuitive, since insight 3 is only a special case of insight 2. Normally distributed demands have 'equal power at all densities', or other words the squared amplitudes A^2 are constant over the whole frequency spectrum (Chatfield, 1996). Applying (10) to i.i.d. normally distributed demand patterns results in:

$$\frac{\text{var } O}{\text{var } I}=\frac{\sum_{i=1}^{N/2-1} A_i^2 AR_i^2}{\sum_{i=1}^{N/2-1} A_i^2}=\frac{A^2 \sum_{i=1}^{N/2-1} AR_i^2}{(N/2-1)A^2}=\frac{\sum_{i=1}^{N/2-1} AR_i^2}{(N/2-1)} \tag{16}$$

The expected variance increase for normally distributed demands from equation (16) is nothing other than the expected value of all AR^2 values. Exactly the same value will be obtained by calculating W_N/π. In the case when the Noise Bandwidth is calculated via numerical integration, we have $W_N=\left(\sum_{i=1}^{N} AR_i^2\right)*\frac{\pi}{N}$. Thus, W_N/π equals the expected value of all AR^2 as well. This result is intuitively clear since in a normally distributed demand pattern, all frequencies are equally present consequently the weights are identical. This result will be quite familiar to control engineers (see e.g. Chatfield, 1996), since "for an input of white random noise, W_N is a direct measure of the variance at the output from the filter" (Towill, 1999). Instead of the term 'white random noise', we use the term "i.i.d. normally distributed demands", but the same conclusion can be made, namely that the Noise Bandwidth is a direct measure for the variance amplification induced by the particular replenishment rule under consideration. For our order-up-to policy defined by equations (1-3) with $T_a=8$ and T_p = 3, W_N/π equals 2.43, and this is the expected increase of the variance for normally distributed demands.

It is interesting to note that, for our order-up-to model with exponential smoothing, the noise bandwidth may be calculated analytically as the integral given by (9) is tractable as shown in (17)

$$\frac{\text{var}(O)}{\text{var}(I)}=\frac{1}{\pi}\int_0^{\pi}\left|\frac{-(2+T_p+T_a)+(3+T_p+T_a)e^{i\omega}}{T_a(-1+e^{i\omega})+e^{i\omega}}\right|^2 d\omega=\frac{13+2T_a^{\;2}+2T_p(5+T_p)+T_a(11+4T_p)}{(1+T_a)(1+2T_a)} \tag{17}$$

This analytical bullwhip measure can be shown, after the necessary replacement of variables, to be exactly the same as the bound found by Chen et al (2000a). This verifies that obviously our z-transform model of order-up-to policies with exponential smoothing is the same as the tight bound on variance amplification given by Chen et al. (2000a), shown below:

$$\frac{\text{var}(O)}{\text{var}(I)}=1+2L\alpha+\frac{2L^2\alpha^2}{2-\alpha}, \text{ with } \alpha=1/(1+T_a) \tag{18}$$

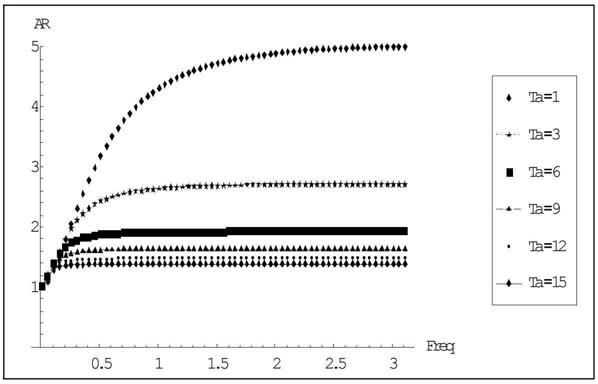
Chen et al. obtained this result for same order-up-to policy (1-3) through statistical analysis. The main conclusion from this analysis is that the statistical approach used by Chen et al. and our control engineering approach result in identical variance amplification predictions. Just for illustrative purpose, the variance amplification equals 4.111 for T_a = 4 and $L=T_p+2 = 5$ and we obtain 2.437 in (17 and 18) for T_a = 8 and 1.677 for T_a = 16.

<u>Insight 4</u>: *Exponential smoothing based order-up-to policies applied to positively* (*negatively*) *correlated demands will result in less* (*more*) *bullwhip effect than when applied to normally distributed demands.*

It can be shown that when applying the FFT to positively correlated demands, the resulting periodogram is a decreasing function of frequency, meaning that the lowest frequencies are most dominantly present. To predict the variance increase, we have to calculate a weighted average of the AR^2 values as explained in insight 2. In this weighted average for positively correlated demands, the weights for the lowest frequencies are larger than the weights for the higher frequencies. And since the AR^2 values for low frequencies are smaller than for high frequencies (see figure (6)), the predicted variance increase will be less than for normally distributed demands. A symmetric argument can be made to prove that negatively correlated demands will result in more bullwhip effect than normally distributed demands. The same conclusion was found by Chen et al. (2000a) using statistical methods.

e. Impact of the smoothing parameter on the bullwhip effect

It is well known that the smoothing parameter has a significant impact on the bullwhip effect. Our control engineering approach confirms this managerial insight. We may highlight this by plotting the FR and computing the Noise Bandwidth for different values of T_a. The results are plotted in the figures 8a and 8b.



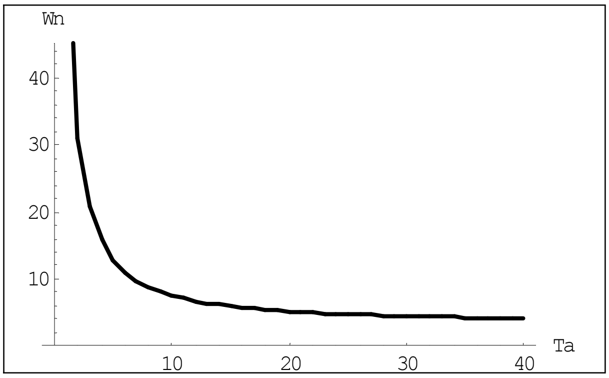


Figure 8a and 8b: Impact of T_a on the bullwhip effect in exponential smoothing order-up-to polices

We observe that the bullwhip effect increases as T_a decreases (and thus α increases). Note that when T_a goes to infinity (i.e. $\alpha \to 0$), we have a fixed order-up-to level for all periods and hence no bullwhip effect. However, large T_a-values are only usable for stable demand patterns. When demand is unstable and irregular and certainly when demand has a trend, small T_a-values have to be chosen to follow the demand closely and obtain a sufficient service level (see Dejonckheere et al (2002) for more details on tracking abilities of common forecasting algorithms). From figure 8 we can see that in those cases the bullwhip effect will certainly be present. Hence there is a trade-off to be made between being responsive and following the demand changes very closely (small T_a - values) on the one hand and avoiding bullwhip (large T_a - values) on the other hand.

4. Order-up-to policies based on other forecasting techniques

In this section, we will use order-up-to policies whose order-up-to levels S_t will be updated by means of moving average forecasting, and 'demand signal processing' (Lee, 1997a, 2000). We will conclude that whatever forecasting method is used, order-up-to policies will always result in a bullwhip effect. The four insights obtained in section 3 will hold under the new assumptions as well.

a. Order-up-to policies with moving average forecasts

We still use the order-up-to policy described in equations (1-2), but now with moving average forecasts used to update the order-up-to levels S_t. The demand forecast of period t, $\hat{D}_t$, is defined as

$$\hat{D}_t = \frac{\sum_{i=0}^{Tm-1} D_{t-i}}{T_m}, \tag{19}$$

with T_m being the number of periods used to compute the forecast. It can be shown (see Appendix B and Wikner, 1994) that the transfer function equals:

$$\frac{O}{D} = \frac{2 + T_p + T_m - (2 + T_p)z^{-T_m}}{T_m} \tag{20}$$

In order to make a fair comparison between exponentially smoothed forecasts with forecasts based on moving averages we have to set (e.g. see Pyke, 1999):

$$\alpha = 2/(T_m + 1) \text{ or } T_m = 2T_a + 1, \tag{21}$$

Applying equation (21), the average age of demand data used in both forecasting systems is identical. In section 3, we used T_a = 8, consequently we use T_m = 17 in this section. The FR plot for T_m = 17 is given in figure 9:

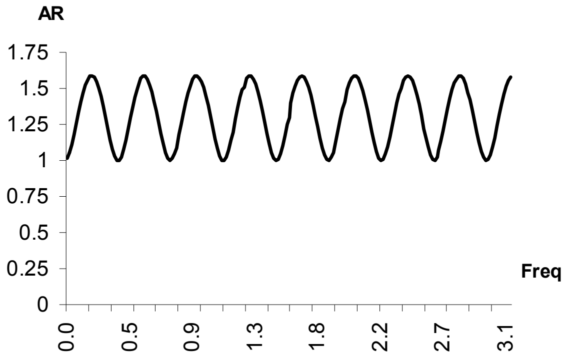


Figure 9: FR for the order-up-to policy with moving average forecasts

Observe the (sinusoidal) shape of the curve: only for a few frequencies, the maximal overshoot (1.588) is obtained. On the other hand, there do exist some frequencies where the overshoot is zero and consequently, there is no bullwhip effect. For sine waves of those particular frequencies, the moving average turns out to be a constant, and thus the order-up-to level S_t is constant, and hence the generated orders are equivalent to the sinusoidal input. Other frequencies will lead to intermediate overshoots. Note that the maximum overshoot is exactly equal to the maximum of the FR curve for the exponential smoothing forecasts (see figure 6). The bullwhip generated by moving average forecasting in order-up-to model therefore is much less than that generated by exponential forecasts.

The variance amplification estimation procedure explained in section 3 and given by equation (15) can easily be repeated for this case. We do however have to use the A^2-values obtained in figure 9. The verification is again based on our spreadsheet simulation. The results for 30 data sets are given in Table I, column 2. For i.i.d. normally distributed demands we can again compare the control engineering based variance amplification metric (Noise Bandwidth / π) with the statistical bound obtained by Chen et al. (2000b) given by:

$$\frac{\mathrm{var}(O)}{\mathrm{var}(I)} = 1 + \frac{2L}{T_m} + \frac{2L^2}{T_m^2} \tag{22}$$

It is satisfying to observe that Chen's statistical approach and our control theory approach result in the exact same outcomes. For T_m = 17 we estimate an increase in variance by a factor of 1.761. The amplification equals 2.728 for T_m = 9 and 1.348 for T_m = 33. However, the control theory route to the analytic expression for bullwhip via the area under the squared frequency plot involves a transcendental equation and the integral cannot be found. The statistical route taken by Chen et al (2000a) has obviously managed to avoid this problem.

It is also important to note that as $T_m=2T_a+1$, these three bullwhip results (for T_m = 9, 17 and 33) may be directly compared to the exponential smoothing results highlighted earlier. Based on these numbers we can conclude that the magnitude of variance amplification is less for moving average based forecasts than for exponentially smoothed forecasts (for an identical age of the demand data used in the forecast). In any case, the bullwhip effect is again guaranteed.

b. Demand signal processing

Consider the following inventory policy:

$S_t = S_{t-1} + \gamma\,(D_t - D_{t-1})$, and $O_t = S_t -$ *inventory position$_t$*, (23)

where O_t is the ordering decision made at the end of period *t*, S_t (S_{t-1}) is the order-up-to level at the end of period *t* (*t*-1), D_t (D_{t-1}) is the observed demand during period *t* (*t*-1) and γ is the 'signalling factor', which is a constant between zero and one. We still have an order-up-to policy, but the order-up-to level is updated every period using the most recently observed demand information. Policies of this type are called 'demand signal processing' by Lee et al. (1997a). For γ = 1, (23) is quite an intuitive policy, often used by human schedulers in real supply chains, Lee et al (1997a). If the retailer experiences a surge of demand in one period, it will be interpreted as a signal of high future demand and a larger order will be placed.

The demand signal processing causal loop diagram is given in Appendix A. For the block diagram, we refer to appendix B. The transfer function for policy (23) is given by:

$$\frac{O}{D} = 1 + \gamma - \frac{\gamma}{z} \qquad (24)$$

The FR plot is given in figure 10 for $\gamma = 1$.

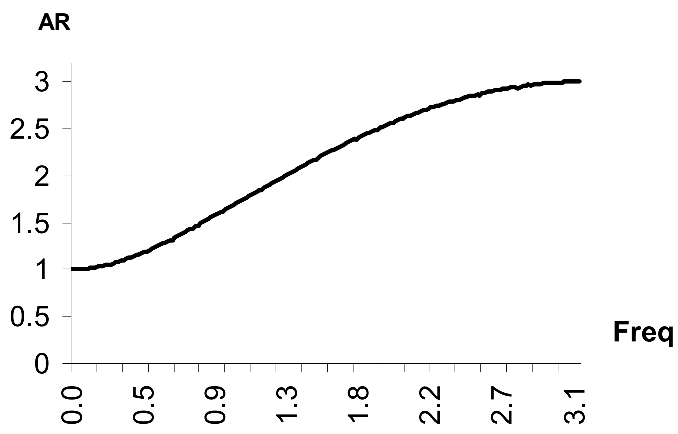


Figure 10: FR for demand signal processing with $\gamma = 1$

The shape of the FR is particularly interesting: there is an overshoot for all different frequencies and the overshoot increases proportionally with frequency. This is intuitively clear since we only use the two most recent demand observations and these short-run demand fluctuations correspond to high frequency signals.

The same insights as explained in section 3 hold for a demand signal processing behaviour. Firstly, there will be a bullwhip effect for every possible demand pattern. Second, the magnitude of the effect can be accurately estimated using equation (15). See Table I, column 3. Third, for i.i.d. normally distributed demand patterns we can easily compute the variance amplification by using (16). For $\gamma = 1$, we obtain an amplification factor of 5. And finally, positively (negatively) correlated demands will result in less (more) amplification than normally distributed demands.

It is interesting to focus on the extremely high amplification factor for $\gamma = 1$. Demand signal processing clearly overreacts dramatically. The overreaction however can be dampened by lowering the adjustment factor γ in equation (23). For $\gamma = 1$, the variance amplification equals 5, for $\gamma = 0.6$, it is reduced to 2.91 and for $\gamma = 0.2$, the amplification equals 1.479. Fractional adjustments clearly result in less amplification.

The integral from w=0 to π of $1+\gamma-\frac{\gamma}{z}$ when $z=e^{iw}$ is tractable (helped by the fact that γ is obviously a real number), thus exploiting the fact that bullwhip=W_N/π for i.i.d demands allows us to generate an analytic expression (equation 25) of bullwhip for order-up-to policies with demand signalling.

$$\frac{\mathrm{var}(O)}{\mathrm{var}(I)} = 1 + 2\gamma(1+\gamma) \tag{25}$$

c. Impact of the insights on the "optimal" use of forecasting within order-up-to policies

The managerial insights presented in this section are very general and have a number of implications on the OR communitity. For instance, given that an optimum forecasting mechanism (whatever that may be) is selected for a particular demand pattern, and that the smoothing parameter(s) is(are) selected to minimise an error function between the demand and the forecast, we have shown that within an order-up-to policy it will always create bullwhip (Insight 1). This is due to the fact the forecasts errors are not the correct focus for the optimisation routine. Analysis of forecasting mechanisms needs to be considered within the context of the entire production and inventory control system. Hence, order rate variations should be analysed if the focus is to minimise bullwhip.

Of course, the exact amount of bullwhip generated by the optimal forecasting system will depend in the forecasting mechanism used and the actual demand signal. The procedure highlighted to prove Insight 2 may be used with confidence to predict actual bullwhip generated by optimal forecasting mechanisms and settings.

In the next section we will design a replenishment rule that will be able to generate smooth ordering patterns, even when forecasting is necessary.

5. A general replenishment rule generating smooth ordering patterns

Order-up-to settings seem to unavoidably result in a bullwhip effect when demand has to be forecasted. In this section we present a general decision rule that does not have that drawback.

a. The decision rule

The order quantity in period t, O_t, is given by:

$$O_t = \hat{D}_t^{T_a} + \frac{1}{T_n}(TNS_t - NS_t) + \frac{1}{T_w}(DWIP_t - WIP_t), \tag{26}$$

where $\hat{D}_t^{T_a}$ is the demand forecast using simple exponential smoothing with parameter T_a, TNS_t a target net stock level, NS_t is the current net stock in period t, $DWIP_t$ is the desired WIP level, and WIP_t finally is the current work in process (or on-order) position in period t. TNS_t is the target net stock level, similar to the safety stock in order-up-to policies. It is updated every period according to the new demand forecast and equals $\hat{D}_t^{T_a}$. $DWIP_t$ is updated every period as well, $DWIP_t = T_p \hat{D}_t^{T_a}$. Note that we only have Tp orders in WIP. T_a, T_n and T_w are the key parameters or controllers of the decision rule. The policy can be described in words as 'ordering quantities are set equal to the sum of forecasted demand, a fraction ($1/T_n$) of the discrepancy of finished goods net stock, and a fraction ($1/T_w$) of our on-order position discrepancy.' The decision rule of equation (26) and small variations of this rule have been described by Towill (1982), John et al. (1994) and Disney (2001). Analysing this replenishment rule from a control engineering perspective offers powerful insights into the variance amplification issue.

b. Relationship with order-up-to policies

Before we derive the transfer function, it is important to see the difference between our policy defined by (26) and an order-up-to policy. The order-up-to policy is defined as follows:

$$O_t = \hat{D}_t^L + k\,\hat{\sigma}_t^L - inventory\ position_t \tag{27}$$

For simplicity, we set $k = 0$ and increase the lead time L by one period. Inventory position equals net stock (NS) + products on order (WIP). We then successively obtain:

$$\left.\begin{array}{l} O_t = (L+1)\,\hat{D}_t^{\,T_a} - NS_t - WIP_t \\ O_t = \hat{D}_t^{\,T_a} + (\hat{D}_t^{\,T_a} - NS_t) + (T_p * \hat{D}_t^{\,T_a} - WIP_t) \\ O_t = \hat{D}_t^{\,T_a} + (\hat{D}_t^{\,T_a} - NS_t) + (DWIP_t - WIP_t) \end{array}\right\} \quad (28)$$

Thus equation (28) turns out to be the complete analogue to the smoothing rule presented in (26) with parameters $T_n = T_w = 1$. In an 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 the key difference with our decision rule of equation (26) in which the errors are included only fractionally. These fractional adjustments are second nature to control engineers, (Towill and Yoon, 1982). It is the reason why the decision rule (26) will be able to generate smooth ordering patterns. Another difference is that in our smoothing decision rule, we have two separate feedback loops (one for the net stock and one for the WIP), whereas in an order-up-to policy, there is only one joint feedback loop for the inventory position. At first sight, these are small differences, but the impact is dramatic.

Note that the decision rule presented in equation (26) is a very general rule. For this paper, we use exponential smoothing to forecast demand, but it is obvious that other forecasting methods can be used. The reader will readily observe that order-up-to policies are actually a special case of our general rule, namely the case $T_n = T_w = 1$.

c. Deriving the transfer function and drawing the frequency response plot

The causal loop diagram for (26) is given below.

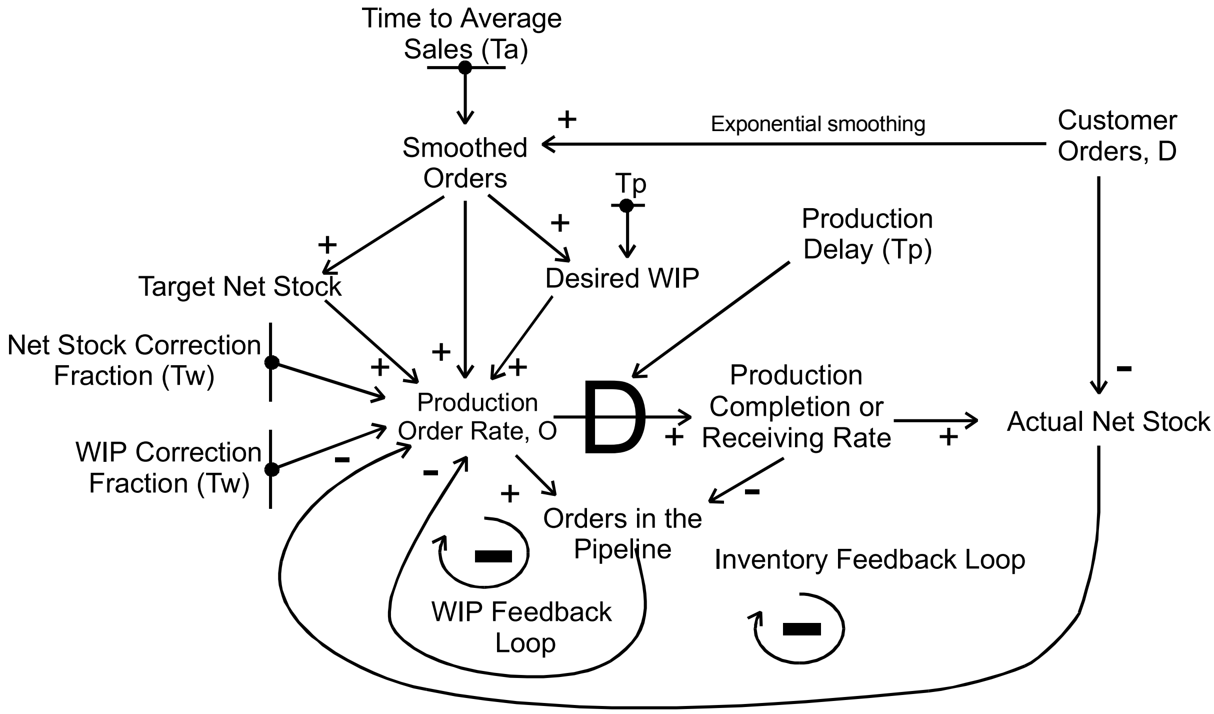


Figure 11: Causal loop diagram for the smoothing decision rule (26)

For the block diagram derivation we refer to Appendix B. Reducing the block diagram yields the following transfer function:

$$F(z)=\frac{O}{D}=\frac{z^{1+Tp}(-(1+T_a)T_w+T_n(T_p+T_w)(-1+z)+(2+T_a)T_w z)}{(T_a(-1+z)+z)(T_w+T_n(-1+(1+T_w(-1+z))z^{T_p}))} \tag{29}$$

Note that if we set $T_n = T_w = 1$ in equation (29), then we obtain transfer function (4). This proves again that order-up-to policies are a special case of the decision rule (26).

Once the transfer function is derived, it is easy to draw the FR plot. For illustrative purposes we set T_a=8, T_n=4 and T_w=4 (here after, denoted as parameter setting (8/4/4)):

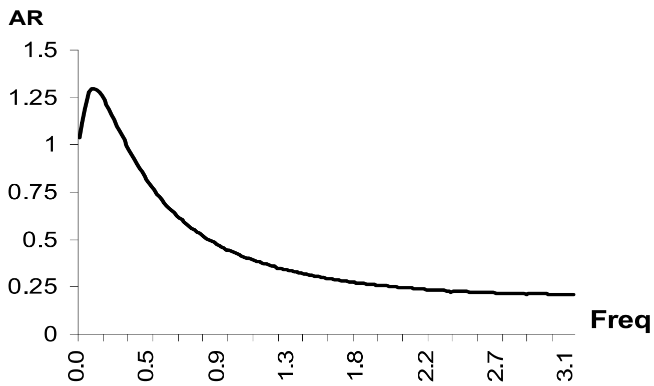


Figure 12: FR for the general replenishment smoothing rule with parameter setting (8/4/4)

We observe a very special and interesting shape of the FR curve in figure 12: there is only overshooting for very few frequencies (the lowest frequencies up to .5 radians per sample interval). For the remaining of the frequency spectrum, the amplitude ratio is actually smaller than one! This gives us the key to bullwhip reduction.

d. Insights for supply chains

Based on the FR plot and spectral analysis we will again enumerate the four insights. Numerical results are given for a (8/4/4) parameter setting. First, replenishment rule (26) is able to reduce variability. There is only overshooting for very few frequencies, and the maximum of the FR plot in figure 12 is less than the peak for order-up-to policies. More importantly, the AR is less than one for the remaining frequencies. This is a very desirable result, because in practice, a decision-making process should be able to correctly identify and track genuine changes in demand (low frequencies). At the same time, the process is expected to detect and reject rogue variations in demand (high frequencies) so that excess costs due to unnecessary ramping up and down production or ordering levels are avoided. Secondly, it is possible to quantify the amount of variability reduction by means of the same procedure explained in section 3. This is shown in column 4 of Table I. For the demand pattern used throughout the text (shipments in figure 1b), we plot the generated smooth ordering pattern in figure 13. Note that the variance ratio is now approximately 0.75 clearly indicating the dampening effect.

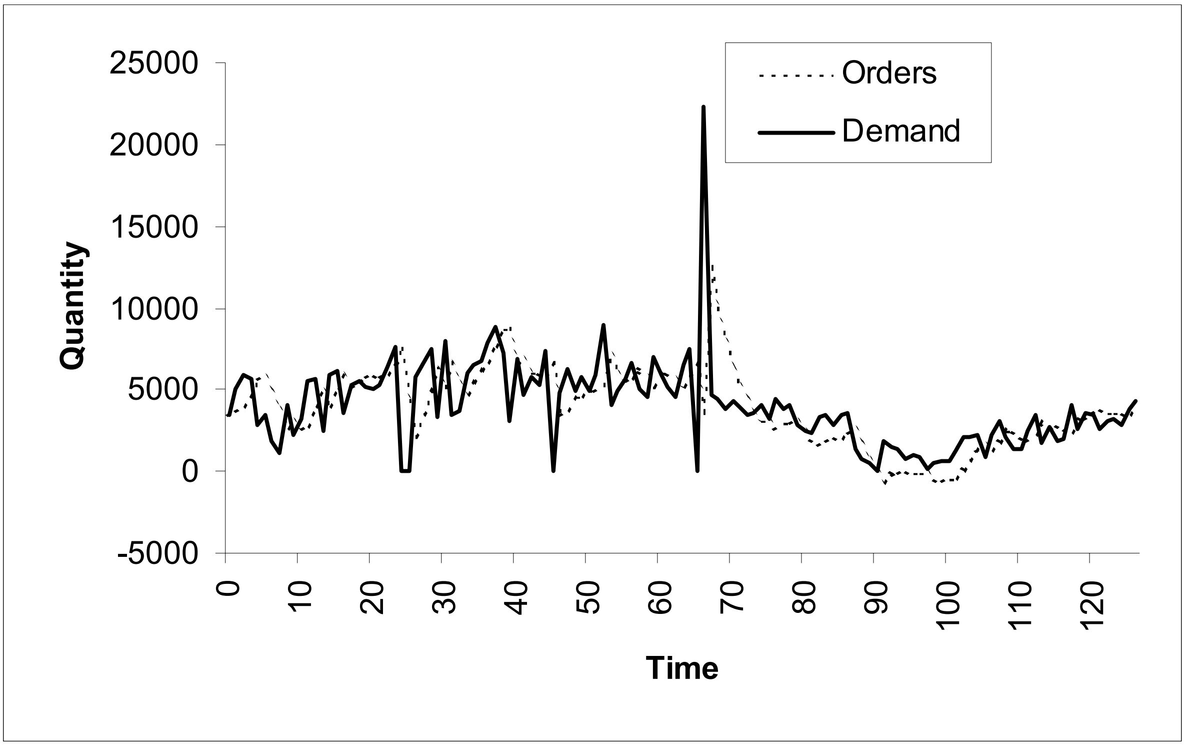
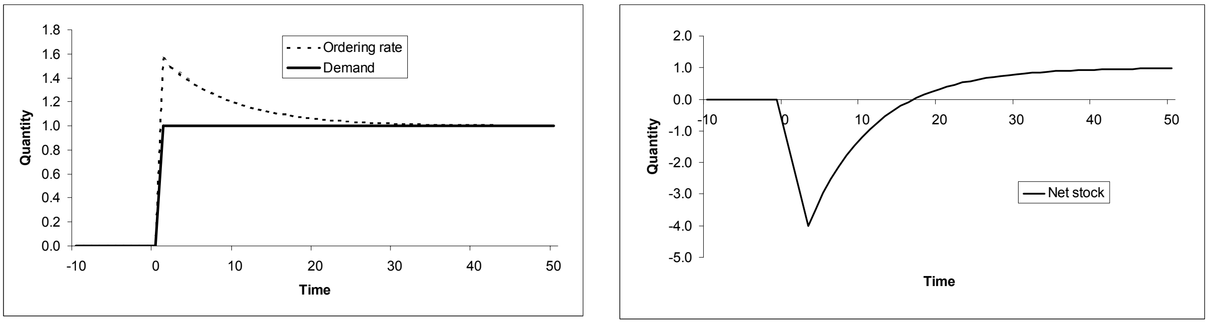


Figure 13: Impact of the smoothing replenishment rule on a real-life demand pattern

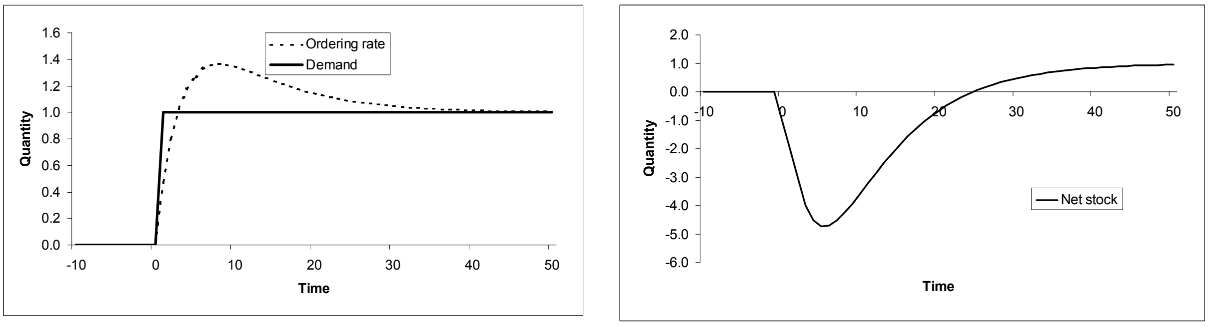
A third observation is that when the rule of equation (26) is applied to normally distributed demands, the ratio of the variances is down to 0.422 (Noise Bandwidth/π). In a fourth observation, we focus on correlated demands. Since low frequencies are most dominantly present in positively correlated demand patterns, and because of the fact that the FR plot shows some restricted overshooting at low frequencies, it goes without saying that the variability reduction for positively correlated demands will be less than the reduction for normally distributed demand patterns. Furthermore negatively correlated demands will be damped substantially more than normally distributed demands.

e. Comparing order-up-to policies with the smoothing decision rule in terms of costs

Let us now compare the order-up-to policy (1-3) with T_a=8 (section 3) with the general smoothing policy (26) with parameter setting (8/4/4). We already know that the former will create variance amplification, whereas the latter succeeds in generating smooth ordering patterns. Hence, the production switching costs will be much larger for order-up-to policies. However, it is important to realise that the order-up-to policy is more responsive to changes in the demand pattern than the smoothing policy when the same smoothing constant T_a is used. To illustrate that in control systems engineering terms, we present both replenishment rules with a ‘step’ input signal, which represents a one time abrupt change in the level of demand. In figure 14, we plot the generated orders following a step input, as well as the resulting changes in the net stock for both policies. We can observe that for the smoothing policy, there is less overshooting in the generated orders compared to the order-up-to policy, but it takes considerably longer for the net stock to recover completely from this step input signal.



Order-up-to policy with exponential smoothing



General replenishment rule with fractional adjustments

Figure 14: Step responses for the order-up-to replenishment rule described by equation (1-3) with T_a=8 (upper graphs) and for the general replenishment rule described by equation (26) with parameter setting (8/4/4) (lower graphs)

As a result of this, the inventory related costs will be larger for the smoothing policy than they were for the order-up-to policy. This is no surprise since order-up-to policies are known to minimise inventory and shortage costs. Hence, there is a trade-off to be made between minimising inventory holding and shortage costs on the one hand and production switching costs on the other side. The choice will be determined by the cost structure of the supply chain under consideration. In this context Elmaghraby (1966) presents an inspirational outline of how to design a replenishment rules to match the cost structure of a supply chain via variance ratios of orders and inventory responses, although this is outside the scope of this particular paper.

It is clear that the selection of the parameters T_a, T_n and T_w will determine the inventory holding and shortage costs as well as the degree of variability reduction that takes place. In figure 12, it can be seen that frequencies up to 0.5 radians per sample interval are 'followed', while frequencies from 0.5 up to π radians per sample interval are 'filtered out'. Other parameters can then be found that ensure more variability reduction takes place. This would again be at the expense of more inventory and holding costs. The choice between signal tracking and signal rejection is well known to control engineers and is described as filter theory. Towill and del Vecchio (1995) have exploited this technique in a supply chain setting.

6. Summary of the replenishment rules explained in the paper

In Table II below, we give an overview of the four replenishment rules (RR) analysed in this paper (column 1): order-up-to policy with exponentially smoothed forecasts, order-up-to policy with moving average forecasting, order-up-to policy with demand signal processing, and the smoothing replenishment rule. For each policy we give the following information: the frequency response plot (column 1) and the presence or absence of the bullwhip effect (Insight 1), the variance ratio for the real life shipments of figure 1b (Insight 2) and for i.i.d. normally distributed demands (Insight 3), and finally the impact of correlated demands on the bullwhip effect (Insight 4).

CONCLUSION

In this paper, we have analysed the bullwhip effect induced by forecasting algorithms in order-up-to policies and we suggest a new general replenishment rule that can reduce variance amplification significantly. Different forecasting methods have been integrated into the order-up-to system. We prove that whatever forecasting method is used (simple exponential smoothing, moving averages or demand signal processing), ***order-up-to systems will always result in the bullwhip effect***. In order-up-to systems, the bullwhip phenomenon is unavoidable when forecasting is necessary; it is the price to pay to forecast unstable demand and to detect trends. Switching production levels up and down frequently may be very expensive in practice. In those cases, it may be important to avoid variance amplification or even to reduce variability of customer demand. We therefore have to design new replenishment rules. We propose a general replenishment rule capable smoothing ordering patterns, even when demand has to be forecasted. The crucial difference with the class of order-up-to policies is that in our proposed rule, net stock and on order inventory discrepancies are only fractionally taken into account. We show that an order-up-to setting is a special case of our general rule.

The methodology is based on control engineering insights. We derive the transfer function and the frequency response plot (FR) for all replenishment rules in the paper. The demand signals are analysed with the Fast Fourier Transform (FFT) in spreadsheets. Combining those techniques, several interesting insights can be found concerning the

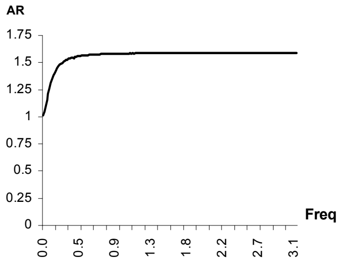
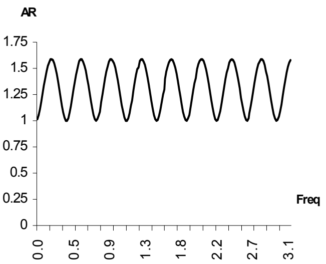
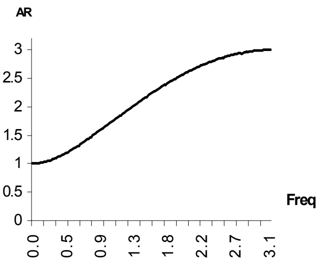
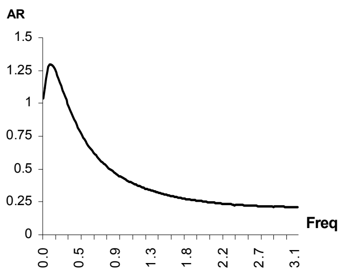
Policy and Frequency Response Plot	Insight No.	Operational Insight Obtained
. 1. Order-up-to with exponential smoothing forecasting  T_a=8	1	Bullwhip is always generated by this policy. Minimum bullwhip of unity when $T_a=\infty$.
	2	Predicted bullwhip from FR = 2.0363 for demand signal shown by shipments in Figure 1b. Actual simulated bullwhip = 2.047.
	3	For i.i.d. demand bullwhip= $W_N/\pi = \dfrac{13+2T_a^{\,2}+2T_p(5+T_p)+T_a(11+4T_p)}{(1+T_a)(1+2T_a)}$
	4	Positively correlated demands create less bullwhip. Negatively correlated demands create more bullwhip.
2. Order-up-to with moving average forecasting  T_m=17	1	Bullwhip is always generated by this policy. Minimum bullwhip of unity when $T_a=\infty$.
	2	Predicted bullwhip from FR = 1.5790 for demand signal shown by shipments in Figure 1b. Actual simulated bullwhip = 1.6089.
	3	For i.i.d. demand bullwhip= $W_N/\pi = 1+\dfrac{2(2+T_p)(2+T_m+T_p)}{T_m^{\,2}}$.
3. Order-up-to with demand signalling forecasting  $\gamma = 1$	1	Bullwhip is always generated by this policy. Minimum bullwhip of unity when $T_a=\infty$.
	2	Predicted bullwhip from FR = 1.5790 for demand signal shown by shipments in Figure 1b. Actual simulated bullwhip = 1.6089.
	3	For i.i.d. demand bullwhip= $W_N/\pi = 1+2\gamma(1+\gamma)$.
	4	Positively correlated demands create less bullwhip. Negatively correlated demands create more bullwhip.
4. Smoothing replenishment rule  T_a=8, T_n=4, T_w=4	1	It is possible to eliminate variance amplification or the bullwhip effect by using fraction adjustments in the inventory and WIP feedback paths.
	2	Predicted bullwhip from FR = 0.7534 for demand signal shown by shipments in Figure 1b. Actual simulated bullwhip = 0.7382.
	3	For i.i.d. demand bullwhip= W_N/π. When Ta=8, Tn=4,Tw=4, Tp=3, bullwhip= W_N/π=0.422.
	4	Positively correlated demands create less smoothing. Negatively correlated demands create more smoothing.

Table II: Summary of the four replenishment rules (RR) analysed in the paper

dynamic behaviour of the replenishment rules and more specifically, we are able to predict whether or not and to what extent they result in variance amplification or smoothing. The presented methodology is very general for two reasons. Firstly, it can be used to analyse the behaviour of every possible replenishment rule for which the transfer function can be derived. In this paper, we mainly focused on order-up-to policies because they are very popular both in supply chain practice and in recent supply chain literature. Secondly, both variance amplification (or dampening) predictions can be made for any possible demand pattern, including real-life data. For the special case of order-up-to policies applied to i.i.d. normally distributed demand patterns, our predictions are identical to the statistically calculated predictions available in the literature.

Acknowledgements

The research contribution from the K.U. Leuven authors was supported by the Science Foundation of Flanders (FWO-project 6.0063.98). We would also like to thank the referees for their careful and detailed reviews of this paper.

Appendix A: Causal loop diagram for 'demand signal processing'

Before drawing the causal loop diagram, we will rewrite equation (20). Recall that the inventory position is net stock plus *WIP*.

$$\left.\begin{aligned} O_t &= S_t - \text{inventory position}_t \\ &= S_{t-1} + \gamma\ (D_t - D_{t-1}) - (\text{inventory position}_t) \\ &= S_{t-2} + \gamma\ (D_{t-1} - D_{t-2}) + \gamma\ (D_t - D_{t-1})\ - (\text{inventory position}_t) \\ &= S_{t-2} + \gamma\ (D_t - D_{t-2})\ - (\text{inventory position}_t) \\ &\ldots.. \\ &= S_0 + \gamma\,(D_t - D_0)\ - (\text{inventory position}_t) \end{aligned}\right\} \tag{A.1.}$$

In this policy, the current inventory position is subtracted from a constant target level plus a 'demand correction term', based on the last observed demand D_t. The causal loop diagram is therefore slightly different from the order-up-to policies based on exponential smoothing or moving average forecasts.

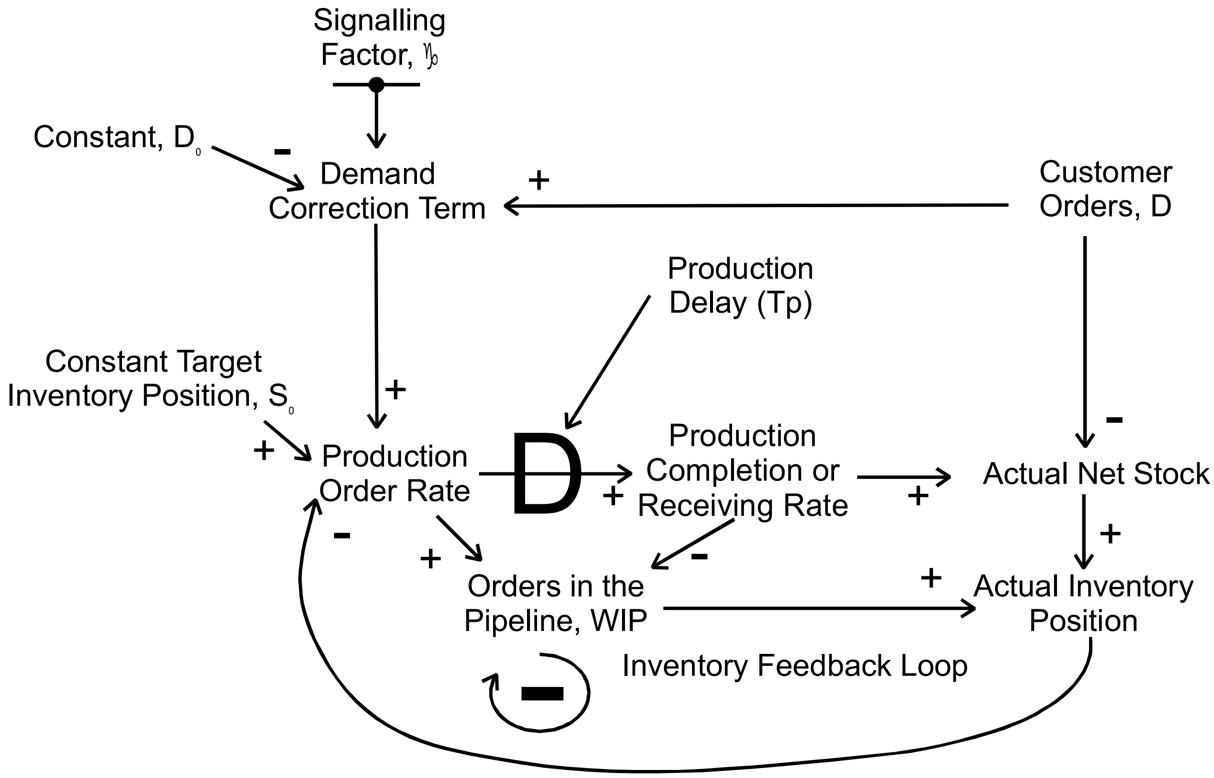


Figure A.1: Causal loop diagram for demand signal processing

Appendix B: A generic block diagram for all the policies treated in the paper

A generic block diagram can be drawn for all policies treated in this paper:

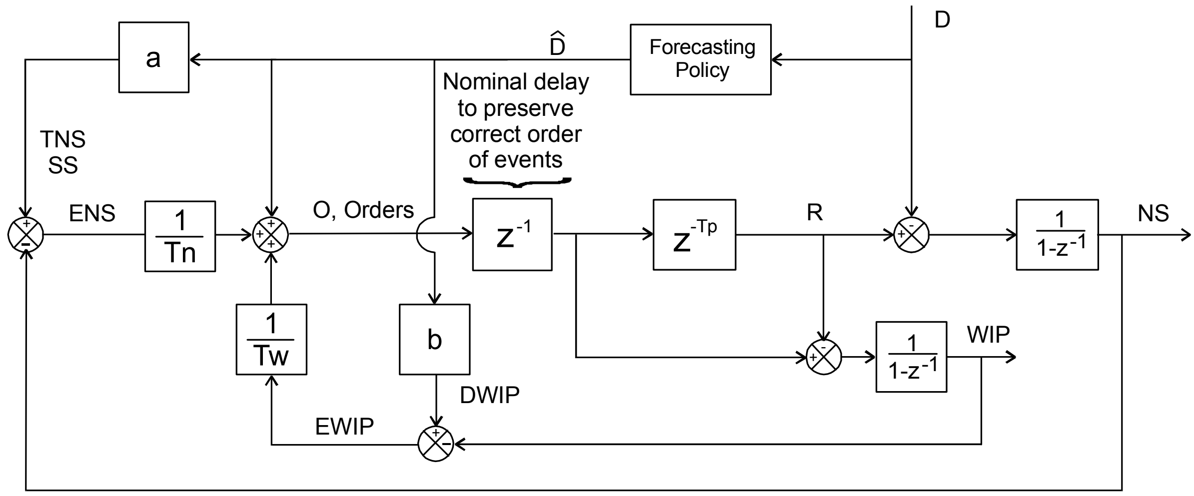


Figure B.1. Generic block diagram for the order-up-to policies studied in the paper

In the block diagram *ENS* stands for the net stock error term and *EWIP* for the WIP error term. The specific parameters for the different replenishment rules are summarised in the table below. Notice that the **a** (the gain between the estimate of average demand and Target Net Stock) and **b** (the gain between the estimate of average demand and Desired WIP) terms in the block diagram need to be substituted for the values in Table B.1. Furthermore the demand signal processing forecasting mechanism reduces to a simple constant.

Inventory Policy	Forecasting Policy	T_n	T_w	Operator a	Operator b
Order-up-to policy with Demand Signal Processing	γ where $0<\gamma\leq 1$	1	1	0	0
Order-up-to policy with Exponential Smoothing	$\frac{\alpha}{1-(1-\alpha)z^{-1}}$; $\alpha=1/(1+T_a)$	1	1	1	T_p
Order-up-to policy with Moving Average	$\frac{1-z^{-T_m}}{T_m(1-\frac{1}{z})}$; $T_m \equiv 2T_a+1$	1	1	1	T_p
Smoothing rule with Exponential Smoothing	$\frac{\alpha}{1-(1-\alpha)z^{-1}}$; $\alpha=1/(1+T_a)$	< 1	< 1	1	T_p

Table B.1. Specific details of individual order-up-to policies shown in figure B.1

Bibliography

Adelson, R.M., 1966. The dynamic behaviour of linear forecasting and scheduling rules, Operational Research Quarterly 17 (4) 447-462.

Aseltine, J.A., 1958, Transform methods in linear system analysis. The Maple Press Company, York.

Baganha, MP, Cohen, MA, 1998. The stabilising effect of inventory in supply chains, Operations Research 46 (3) 572-583.

Bertrand J.W.M., 1986. Balancing production level variations in complex production systems, International Journal of Production Research, 24 (5) 1059-1074.

Bonney, M.C. and Popplewell, K., 1988, Design of production control systems: Choice of systems structure and systems implementation sequence. Engineering Costs and Production Economics, 15, 189-173.

Bonney, M.C., Popplewell, K. and Matoug, M., 1994, Effect of errors and delays in inventory reporting on production system performance. International Journal of Production Research, 35, 93-105.

Chatfield, C., 1996. The analysis of time series: an introduction, Texts in statistical sciences, Chapman&Hall pp283.

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

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

Cochran W.T., Cooley, J.W., Favin, D.L., Helms, H.D., Kaenel, R.A., Lang, W.W., Maling, G.C., Nelson, D.E., Rader, C.M., Welch, P.D., 1967, What is the Fast Fourier Transform?, IEEE Transactions on Audio Electro-acoustics AU-15 (2) 45-55.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R., 2002. Transfer function analysis of forecasting induced bullwhip in supply chains, International Journal of Production Economics, accepted for publication.

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

Disney, S.M., 2001. The production and inventory control problem in Vendor Managed Inventory supply chains, PhD. Thesis, Cardiff University, UK.

Elmaghraby, S.E., 1966, The design of production systems. Reinhold Publishing Corporation, New York.

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

Forrester, J., 1961. Industrial Dynamics, Cambridge MA, MIT Press.

Garnell, P., East, D.J., 1977. Guided weapons control systems, Pergamon Press, Oxford.

Grubbström, R.W., 1998, The fundamental equations of MRP theory in discrete time. Working Paper No. 254, Department of Production Economics, Linköping University, Sweden.

Houpis, C.H., Lamont G.B., 1985, Digital Control Systems. McGraw-Hill, New York.

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

Jury, E.I.,1964, Theory and application of the z-transform method, Robert E. Krieger Publishing Company, New York.

Kaminsky, P., Simchi-Levi, D., 1998. A new computerized beer game: a tool for teaching the value of integrated supply chain management, POMS Series in Technology and Operations Management 1 pp 216-225.

Karlin, S., 1958. One stage inventory models with uncertainty. In: Karlin, S., Scarf, H., Studies in the mathematical theory of Inventory and Production, Stanford University Press, 109-134.

Lambrecht, M., Dejonckheere, J., 1999. Extending the Beer Game to Include Real-Life Supply Chain Characteristics. In: Bartezzaghi E., Filippini R., Spina G., Vinelli A., Managing

Operations Networks, 237-243.

Lee, H.L., Padmanabhan, V., Whang, S., 1997a. The Bullwhip effect in supply chains, Sloan Management Review, Spring 93-102.

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

Lee, H.L., So, K.C., Tang, C.S., 2000. The value of information sharing in a two level supply chain", Management Science, 46 (5) 628-643.

Makridakis, S., Wheelwright, S.C., McGee, V.E., 1978. Forecasting: methods and applications, John Wiley & Sons.

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

Porges, S.W., Bohrer, R.E., 1991. The analysis of periodic processes in psychophysiological research, working paper (downloadable from http://www.wam.umd.edu/~sporges/bohrer.htm)

Silver, E., Pyke, D., Peterson, R., 1998. Inventory management and production planning and scheduling, John Wiley & Sons, pp 754.

Simchi-Levi, D., Kaminsky P., Simchi-Levi, E., 2000. Designing and managing the supply chain, concepts, strategies and case studies, McGraw-Hill.

Simon, H.A., 1952. On the application of servomechanism theory in 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.

Sterman, J., 2000. Business dynamics, systems thinking and modelling for a complex world, Irwin McGraw-Hill, pp 982.

Towill, D.R., 1970. Transfer Function Techniques for Control Engineers, Iliffe Books, London.

Towill, D.R., 1982. Dynamic analysis of an inventory and order based production control system, Int. Jnl. Prod. Res. 20 369-383.

Towill, D.R., 1999. Fundamental theory of bullwhip induced by exponential smoothing algorithm, MASTS Occasional Paper No. 61, Cardiff University.

Towill, D.R., del Vecchio, 1995. Application of filter theory to the study of supply chain dynamics, Production Planning and Control 5 1 82-96.

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

Vassian, H.J., 1955. Application of discrete variable servo theory to inventory control, Operations Research 3 272-282

Wikner, J., 1994. Dynamic modelling and analysis of information flows in production-inventory and supply chain systems, PROFIL series (12), Production-Economic Research in Linköping.