The economic consequences of a production and inventory control policy

Stephen M. Disney
Logistics Systems Dynamics Group, Cardiff Business School, Cardiff University,
Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK. Email: DisneySM@cardiff.ac.uk

And

Robert W. Grubbström
Department of Production Economics, Linköping Institute of Technology,
S-581 83 Linköping, Sweden. Email: rwg@ipe.liu.se

ABSTRACT
We investigate the economic performance of a generalised Order-Up-To policy in response to an Auto Regressive stochastic demand process. We focus on the case where the physical production/distribution lead-time is one period and where we forecast demand with simple exponential smoothing. We consider two sets of convex piece-wise linear costs. The first set is the traditional inventory holding and backlog costs. The second set of costs are piece-wise linear and increasing convex costs associated with the production order rate within and above a capacity constraint. Numerical investigations reveal that the classical Order-Up-To policy is no longer optimal when a broader range of costs is considered in the objective function.

Keywords: Bullwhip, Inventory variance, Order-Up-To Policy, Expected costs

1. INTRODUCTION
There is a large body of inventory related theory that optimises the inventory holding and backlog costs within a single business. However, when these “optimal” policies are strung together in a supply chain, they create the “bullwhip” problem, Kahn (1987) and Lee et al (1997). The bullwhip problem is where the variance of the order signal increases as the order flows up the supply chain. Forrester (1958) showed us that this demand variance amplification problem is caused by the structure of the replenishment decisions used by each echelon in the supply chain as it reacts to their individual demand signals. It has been estimated that the economic consequences of the bullwhip effect can be as much as 30% of factory gate profits, Metters (1997). The negative effects of bullwhip problem have been further summarised by Carlsson and Fullér (2000) as follows;

- Excessive inventory investments throughout the supply chain to cope with the increased demand variability
- Reduced customer service due to the inertia of the production/distribution system
- Lost revenues due to shortages
- Reduced productivity of capital investment
- Increased investment in capacity
- Inefficient use of transport capacity
• Increased missed production schedules

So, we ask ourselves, “Rather than just concentrating on inventory variance, why don’t we bring the cost of the order variance into the evaluation and design of the PIC system?” It is this question we address here.

The ordering policy that we study here is an infinite horizon, discrete time, periodic review Order-Up-To (OUT) policy. That is, at discrete equally spaced moments in time (i.e. every day, week, month), we review our inventory position and order-up-to a suitable amount. Note that we place an order for the product to be produced every planning period (day, week, month etc.) and receive it some time later. Also note that we consider the OUT policy which is also the periodic (s,S) policy when s=S. We concentrate on the case where the physical production/distribution lead-time is one time period (and there is also a review period), although the case for different lead-times follows essentially the same argument.

We also need a demand signal in our analysis. We have chosen to use the weakly stationary stochastic Auto Regressive (AR) demand process motivated by a stochastic variant from the normal distribution, Box and Jenkins (1970). This demand process is strictly stationary (hence we may determine its long-run, unconditional, variance and mean), but it does exhibit some non-stationary characteristics that justify the use of a forecasting mechanism, Disney et al, (2002). We use exponential smoothing, Brown (1962) as a forecasting mechanism within the OUT policy to determine the replenishment orders as did Chen et al (2000). We modify the classical OUT policy to yield an ordering system that has much greater flexibility in the trade-off between bullwhip and inventory variance by incorporating proportional controllers into the two feedback loops as in Dejonckheere et al (2003). The contribution of Dejonckheere et al (2003) was to show for all lead-times and all possible demand patterns the classical OUT policy with exponential smoothing or moving average forecasts always results in bullwhip. They further showed that a proportional controller in the feedback loops allowed this bullwhip effect to be avoided. Herein we focus on the economic consequences of this bullwhip avoidance mechanism.

As we treat time as discrete, we will exploit the z-transform to develop a model of the ordering policy and the demand process in our methodology, Vassian (1955). We consider a linear response to the AR demand process to be possible and hence we may exploit transfer functions and difference equations to model the OUT policy’s response. We use Tsypkin’s (Tsypkin 1964) relation to derive closed form expressions of the variance of the replenishment orders and inventory levels over time directly from the difference equations. We assume that stock-outs are fully backlogged and that there is an alternative source of supply when capacity has been exceeded. Thus, from the mean and variance of inventory and orders we may determine the expected number of products per period that will be produced in normal and expedited (or premium) production modes and the expected inventory holding and backlog per period (and hence their expected costs). We do however, place different values to inventory backlog and holding costs and to normal and expedited production costs. We then highlight the economic consequences of the parameter setting in a simple numerical example. Our methodology is completely analytical and exact.

Our results confirm that the classical OUT policy does indeed minimize the NPV of the inventory related costs when demand is stationary and independently and identically distributed.
(i.i.d.). This is well known and not surprising, Kahn (1987). However, if the objective is to minimise both the inventory and order related costs (i.e. to include the costs associated with the bullwhip problem) then the classical OUT policy is no longer optimal. However, our modified OUT policy is capable of reducing the total NPV of these order and inventory payments. Our modified OUT policy can also do better than the classical OUT policy solely in terms of inventory costs (holding and backlog) when demand is not i.i.d.

We proceed as follows. First we introduce and formally describe the OUT policy in section 2. Section 3 defines the AR demand process. In section 4 we present the variance ratios that describe the inventory levels and production orders over time. Section 5 introduces our cost function and derives expectation expressions of order and inventory positions in each time period. In section 6 we highlight the expected total cost per period by numerical example and pay special attention to 4 common production scheduling strategies. Section 7 concludes.

2. THE ORDER UP TO POLICY

The ordering policy we have chosen for our analysis is a generalized OUT policy. In a classical OUT policy the order is calculated as,

\[ O_t = S_t - \text{inventory position } t \]  

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 inventory on order (or WIP). The Net Stock (NS) equals inventory on hand minus backlog. The order-up-to level is updated every period according to

\[ S_t = \hat{D}^t + k\hat{s}^t, \]  

where \( \hat{D}^t \) is an estimate of mean demand over \( L \) periods (we assume \( \hat{D}^t = L\hat{D}^{t_e} \), where \( \hat{D}^{t_e} \) is the estimate of demand in the next period calculated with exponential smoothing, with smoothing constant \( T_a \)), \( \hat{s}^t \) is an estimate of the standard deviation of the demand over \( L \) periods, and \( k \) is a chosen constant to meet a desired service level. To simplify the analysis many authors, set \( k \) equal to zero and increased the lead-time by one. However, we elect to set \( k \) equal to zero and increase the lead-time by a variable \( \hat{L} \) (where \( \hat{L} \geq 0 \)). This results in a more general form of the OUT model.

Additionally, \( L \) is increased by one to ensure the correct order of events. We essentially follow the order of events due to Vassian (1955). For example, we receive inventory and satisfy demand throughout the planning period and at the end of the planning period we observe inventory 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. In other words \( L \) not only represents the physical lead-time, \( T_p \), but also a safety lead-time (\( \hat{L} \)) and an order of events delay, the so-called review period, (+1). Thus we have \( L = T_p + \hat{L} + 1 \). For simplicity, we assume herein that the physical production / distribution lead-time, \( T_p \), is 1.

Finally the order-up-to policy definition is completed as follows; inventory position equals net
stock (NS) + products on order but not yet received (WIP). Writing $DWIP = Tp \dot{D}_t$, we then successively obtain:

$$
\begin{align*}
O_t &= (Tp + \hat{L} + 1) \dot{D}_t - NS_t - WIP_t, \\
O_t &= \dot{D}_t + (\hat{L} \dot{D}_t - NS_t) + (Tp \dot{D}_t - WIP_t), \\
O_t &= \dot{D}_t + (\hat{L} \dot{D}_t - NS_t) + (DWIP - WIP_t).
\end{align*}
$$

As hinted at earlier, will make the following modification to the OUT policy so that we may avoid the Bullwhip Effect as shown by Dejonckheere et al (2003).

$$
O_t = \dot{D}_t + \frac{\dot{L} \dot{D}_t - NS_t}{Ti} + \frac{DWIP - WIP_t}{Ti},
$$

From the above description, we may draw the following block diagram of the modified OUT policy where $1/Ti$ is the gain of the proportional feedback controller in the inventory and WIP feedback loops and $Ta$ is the exponential smoothing constant in the forecasting mechanism. The valid ranges of these parameters needed to ensure stability are; $0.5 < Ti < \infty$ and $-0.5 < Ta < \infty$.

![Figure 1. Block diagram of the generalised OUT policy](image)

**Figure 1.** Block diagram of the generalised OUT policy

### 2.1 Transfer functions

Manipulating the block diagram using standard techniques (see Nise 1995 for an introduction) for the transfer function of orders results in (5).

$$
\frac{Z[O]}{Z[D]} = \frac{(1 + \hat{L} + Ta + Tp + Ti)z^2 - (\hat{L} + Ta + Tp + Ti)z}{(1 + Ti(-1+z))(Ta(-1+z) + z)}
$$

Similarly the Net Stock transfer function is given by;
3. THE DEMAND PATTERN
We have chosen the AR demand pattern as a suitable demand pattern. The mean centred AR demand pattern may be generated from stationary white noise as follows;

\[
\begin{align*}
D_{i,t} &= \mu_D + \varepsilon_t, \\
D_{t,r} &= \rho(D_{(t-1),r} - \mu_D) + \varepsilon_r + \mu_D,
\end{align*}
\]  

where:

- \(\mu_D\) = the mean of the stochastic demand pattern (which we may set arbitrarily high to effectively eliminate negative demand, i.e. \(\mu_D > 4\sigma_D\)).
- \(\varepsilon_r\) = white noise that is the input into the demand generator. We assume that this is a standard normal distribution with a mean of zero and unit variance.
- \(\rho\) = auto regressive coefficient. -1 < \(\rho\) < 1.
- \(D_{t,r}\) = AR demand at time \(t\).

We may convert the difference equation representation (7) of the demand pattern into a transfer function. This transfer function describes completely the demand pattern in the discrete complex frequency domain;

\[
\frac{Z\{D_{AR}\}}{Z\{\varepsilon\}} = \frac{z}{z - \rho}
\]  

4. THE VARIANCE RATIOS
Our methodology for determining closed form expressions of the unconditional variance of inventory levels and the production order rates over an infinite horizon is shown in Appendix 1 where it is applied to the AR demand process. We will not show other workings here due to the lengthy nature of the algebra involved. In order to assist in the algebraic manipulation during our investigation we exploited Mathematica (Wolfram Research) and verified our work with a difference equation model in Microsoft Excel.

The unconditional long-run variance of the production order rate is given by,

\[
\sigma_\rho^2 = \frac{\left(-(1 + Ta)T(t^4 + 5Ta + 7Ti + 2T^2 + 3TaT^2 + T^2 + \hat{L}(3 + 2Ta + 2Ti))\right) + T(t + Ti)^5 + 7Ta + 5Ti + 2T^2 + 3TaT^2 + Ti^2 + \hat{L}(3 + 2Ta + 2Ti))\rho + Ta(-1 + Ti)^4 + 5Ta + 7Ti + 2T^2 + 3TaT^2 + Ti^2 + \hat{L}(3 + 2Ta + 2Ti))\rho^2}{\left(1 + 2Ta(T(t + Ti)^2 - 1 + 2Ti)\right)\left(-1 + Ta(-1 + \rho)\right)\left(T(t + Ti^2 - 1 + \rho) - \rho\left(-1 + \rho^2\right)\right)}
\]
that we may plot as follows for the case of $\rho=0.9$ for different values of $Ta$ and $Ti$, the introduced gain in the two feedback loops. Note that we have plotted $1/Ti$ as the permissible range of $Ti$ required for a stable response (see Disney et al. (2003)) is $Ti>0.5$, and plotting $1/Ti$ allows the impact of the complete range to be viewed concisely.

![Figure 2. Order variance when $\rho=0.9$, $\hat{L}=0.1$.](image)

Figure 2 shows both the impact of $Ta$ (the average age of the exponentially smoothed forecast of demand used to determine the OUT point $S$) and the feedback loops gain $Ti$. Slower moving forecasts (i.e. larger values of $Ta$) dampen the variance of the order rate as do larger values of $Ti$. The unconditional mean of orders is obviously $\mu_O = \mu_D$.

The unconditional variance of the Net Stock levels over time is given by:

$$
\sigma_{NS}^2 = \frac{\left(1 + Ta Ti\right)\left(Ta(2 + 2 \hat{L} + \hat{E} + 2 Ta) - (2 + 4 \hat{L} + \hat{E} + 2(1 + \hat{L})[2 + \hat{L}] Ta^2)Ti - 2(2 + Ta(4 + Ta))Ti^2\right) + 
\left(Ta + Ti\right)\left(2 \left(2 - 2 Ta(-1 + Ti) + 2(-1 + Ta(-1 + Ti))\right)^2 + \hat{L}(1 + Ta - Ti - 2 Ta Ti)\right)\rho + 
\left(2 Ta(1 + \hat{L} + Ti)(1 + 2 Ti) + 2 Ta\left(1 + Ti - Ti^2 + Ti^3\right)\right) + 
Ta\left(\left(2(1 + \hat{L}) - (6 + \hat{L}(4 + \hat{L}))(1 + 12 Ta Ti) + 12 Ti^2\right) + \rho^2\right) 
\left(1 + Ta Ti(1 + 2 Ta Ti)(-1 + 2 Ti)(-1 + Ta(-1 + Ti))(Ti(-1 + \rho) - \rho)(-1 + \rho^2)\right)}\left(1 + Ta + Ti\right)^2 \left(-1 + 2 Ti\right)\left(-1 + Ta\right)\left(1 + 2 Ti\right)\left(-3 Ta Ti + 2(6 + \hat{L}(3 + \hat{L}))(Ti^2 + 2 Ti^3) - 2(2 + Ti)\right)}
$$

(10)

that we have plotted for the case when $\rho=0.9$ for various $Ta$ and $Ti$ as an illustration.
Figure 3 reveals that the variance of the inventory level for the generalised OUT policy is concave in \( Ti \) and \( Ta \) for the considered values of \( \rho \) and \( \hat{L} \). The value of \( Ti \) that minimises the inventory variance (i.e. where inventory and backlog costs are also obviously minimised) is clearly influenced by the value of \( Ta \). Larger values of \( Ta \) require a smaller value of \( Ti \). This relationship is also influenced by the value of \( \rho \), the auto regressive co-efficient in the AR demand. For example, compare Figure 3 with Figure 4. In Figure 4 we have set \( \rho = 0 \) to create an i.i.d. demand process and \( Ta = \infty \), as this is known to minimise the \( n \) period ahead forecast error when \( \rho = 0 \). We see the optimal \( Ti \) that minimises the inventory variance (and thus minimises inventory holding and backlog costs) is unity. The role of \( Ti \) is symmetrical around \( Ti=1 \) for this i.i.d. demand process.
The mean of the Net Stock levels is obviously $\mu_{NS} = \hat{L}\mu_D$.

5. EXPECTED COSTS PER PERIOD
The variance ratios presented in the previous section hold if $\varepsilon$ is drawn from any i.i.d. random distribution such as normal, log normal, exponential etc, see Grubbström and Andersson (2002). However we will now assume it is a normal distribution and as the Demand, Orders and Net Stock positions are a linear combination of the normal distribution they will also be normally distributed.

Running a simulation of our OUT policy motivated by an actual normal random variable we obtain a typical time series of orders and net stock positions as shown in Figure 5. It is easy to visualize the costs we are going to incorporate into our analysis here. For the normal production rate, we assume a capacity of 12.5 units per planning period. Recall that we are considering a linear response, thus when the orders in each period are above the normal capacity limit and that we assume there is another source of supply (albeit at a premium). Examples of this alternative source of supply may include over-time working, purchasing or subcontracting. In the second graph, we can see a typical Net Stock time series. Clearly some of the time Net Stock is negative and hence we assume the orders are fully backlogged. Our task now is to assign expected costs to the; backlog position, inventory holding position, production completed within normal capacity and the production completed in premium capacity. This is to be done as described by (11 and 12).

$$\text{OrderCosts} = \begin{cases} O.A, O \leq C \\ A.C + F(O-C), O > C \end{cases} \quad (11)$$

$$\text{InventoryCosts} = \begin{cases} G.(-NS), NS \geq 0 \\ H.NS, NS < 0 \end{cases} \quad (12)$$

where:
- $A = \text{unit cost of production when in normal working hours}$
- $F = \text{unit cost in over-capacity production}$
It is also useful to look at the order position over time as a probability density function, see Figure 6. We know the mean and variance of the order position, and we can define the capacity limit, $C$, and costs as is shown in Figure 6. Here $\mu_o$ and $\sigma_o$ is the mean and standard deviation of the order rate respectively.
Figure 6. Visualisation of normal and expediting production costs

We may then build up the following expression for the expected number of units per period produced within the normal capacity limit $C$ as defined by (11) as follows;

$$E[N] = \mu - \frac{1}{\sigma_0 \sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{(C-\mu-x)^2}{2\sigma_0^2}} dx$$

(13)

Similarly the expected number of units associated with the expedited production above the capacity limit per period is given by;

$$E[P] = \frac{1}{\sigma_0 \sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{(C-x)^2}{2\sigma_0^2}} dx$$

(14)

The Net Stock probability function may also be illustrated as shown in Figure 7. Note we have incorporated piece-wise linear costs again.
Inspecting Figure 7, we may write the expected inventory holding and backlog position per period as given by (13) and (14), respectively.

\[ E[I] = \hat{L} \mu + \frac{1}{\sigma_{NS} \sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{(L \mu + x)^2}{2\sigma_{NS}^2}} (-x)dx \]  \hspace{1cm} (15)

\[ E[B] = \frac{1}{\sigma_{NS} \sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{(L \mu + x)^2}{2\sigma_{NS}^2}} (-x)dx \]  \hspace{1cm} (16)

6. NUMERICAL ANALYSIS OF EXPECTED COSTS

Consider the following numerical example; \( \mu_D = 10 \), \( \rho = 0.9 \), \( \hat{L} = 0.1 \), \( C = 12.5 \), \( A = 10 \), \( F = 20 \), \( G = 3 \), \( H = 6 \). The expected Total Costs (TC) are given by;


Enumeration of the expected total costs per period produced by the generalised OUT policy under these settings is shown in Figure 8.
Consider the following observation. In a perfect scenario, all products would be produced in normal production and there would be no inventory holding or backlog costs. In our numerical scenario this would give us expected unavoidable costs per period of 100. This will be used as a benchmark for the comparison we have highlighted in Table 1. Here we consider 4 common production-scheduling strategies. The first is the level scheduling strategy where we set $Ta=Ti=99$ to produce a reasonably level production schedule. In this case there is a high inventory variance and the policy creates avoidable costs of 266.55 per period, on the average. The second strategy, Pass On Orders, where the production order rate is simply the last observed demand does significantly better by reducing avoidable costs down to 16.09 per period. An optimal classical OUT policy, where the exponential smoothing forecast has been tuned to minimum expected costs results in 11.09 units of cost per period. The global minimum of our generalised OUT policy reduces costs further to 11.22. Here we have two minimum total cost scenarios as both the order variance and inventory variance is symmetrical about $Ti=Ta+1$. 

Figure 8. Expected costs incurred by the generalised OUT policy for $Ta$ and $Ti$
Table 1. Economic performance of some common production scheduling strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Ta</th>
<th>Ti</th>
<th>Unnecessary Costs</th>
<th>$\sigma_{NS}^2$</th>
<th>$\sigma_{O}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure level scheduling</td>
<td>99</td>
<td>99</td>
<td>166.556</td>
<td>2189</td>
<td>1.11057</td>
</tr>
<tr>
<td>Pass on orders</td>
<td>99</td>
<td>1</td>
<td>16.086</td>
<td>18.5556</td>
<td>5.4681</td>
</tr>
<tr>
<td>Optimal classical OUT</td>
<td>0.873852</td>
<td>1</td>
<td>11.281</td>
<td>5.90413</td>
<td>8.84972</td>
</tr>
<tr>
<td>Global minimum in the generalised OUT policy</td>
<td>-0.18374</td>
<td>1.46997</td>
<td>11.216</td>
<td>5.85532</td>
<td>8.78238</td>
</tr>
</tbody>
</table>

7. CONCLUSIONS
We have analysed the economical impact of order and inventory related cash flows resulting from a generalised OUT policy. We have used $z$-transforms and probability density functions to obtain exact results. The complete solution space available is clearly very large and we have only considered part of it herein. However, we have shown that our modification to the OUT policy (that is incorporating proportional controllers in the two feedback loops) is economically desirable for a particular scenario and a particular set of cost functions. Clearly more research needs to be done in this area. Of particular interest here is the “Axsäter integrated production-inventory system”, where individual machines and multiple products and components may be considered in a matrix framework as discussed in Grubbström and Lundquist (1977).

Acknowledgments: We would like to thank Professor Denis Towill (Cardiff University, Wales), Professor Marc Lambrecht (Leuven University, Belgium), Professor Frank Chen (Chinese University of Hong Kong), Professor Mohamed Naim (Cardiff University), Wim van de Velde and Ingrid Farasyn (Procter and Gamble, Belgium) and Jeroen Dejoneckheere (GE Control Systems, Belgium) who helped to formulate our understanding of some of the techniques exploited in this paper. We would also like to thank Cardiff University's Young Researcher Initiative for sponsoring this research.

8. APPENDIX A: DERIVING THE VARIANCE RATIOS
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 $AR$ 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 (7) of the AR demand generator we first convert it into a $z$-transform model via the block diagram shown in Figure A1.

![Figure A1. Block diagram of the AR demand pattern generator.](image-url)
Manipulation of this block diagram using standard techniques yields the transfer function given by (8) (where $z$ is the $z$-transform operator) that describes completely the demand pattern in the discrete complex frequency domain. In order to calculate the variance amplification ratio between the pure white noise input and the AR demand pattern we exploit Tsypkin’s relation (Tsypkin, 1964) that states that the variance of a systems output divided by the variance of the input (when subject to an input of pure white noise) is equal to the sum of the squared impulse response in the time domain. So we take in the inverse $z$-transform of (8) to find the time domain impulse response,

$$\frac{D_{AR}(n)}{\varepsilon(n)} = \rho^n$$  \hspace{1cm} (A1)

and then sum the square from zero to infinity to find the variance ratio between the white noise input and the AR demand.

$$\frac{\sigma_{D_{AR}}^2}{\sigma_{\varepsilon}^2} = \sum_{n=0}^{\infty} (\rho^n)^2 = \frac{1}{1 - \rho^2}$$  \hspace{1cm} (A2)

We have used this technique throughout this paper, without referring to or presenting the details, as the equations involved are often very lengthy. An alternative method, staying in the $z$-domain, is provided in Grubbström and Andersson, 2002.

9. REFERENCES


