Monograph:
Riddalls, C.E., Bennett, S. and Tipi, N.S. (1999) The Effect of Batched Production on Demand Amplification. Research Report. ACSE Research Report 753 . Department of Automatic Control and Systems Engineering

X

Research Report No. 753

The Effect of Batched Production on Demand Amplification*

C.E. Riddalls, S. Bennett and N.S. Tipi

Department of Automatic Control and Systems Engineering, Universty of Sheffield,

S1 3JD. Tel 0114 222 5186, fax 0114 273 1729, c.e.riddalls@sheffield.ac.uk

Demand amplification is the tendency of small fluctuations in demand at the retailer end of the supply chain to be amplified as they are communicated down the chain. A brief review of the literature on this phenomenon is presented, concentrating particularly on the causes propounded. A continuous-time differential equation model of a production-inventory system is then proposed. The application of a novel optimal control algorithm is applied in order to simulate the rational behaviour of inventory managers. This algorithm allows us to mimic the discontinuous cost structures implied by the advantages of batched production. By simulating the response of the system to small changes in demand, the relationship between batch size and the magnitude of demand amplification is investigated.

August 1999

* This work was supported by EPSRC under the grant number GR/M27432

The Effect of Batched Production on Demand Amplification*

C.E. Riddalls, S. Bennett and N.S. Tipi

Department of Automatic Control and Systems Engineering, University of Sheffield, S1 3JD. Tel 0114 222 5186, fax 0114 273 1729, c.e.riddalls@sheffield.ac.uk

Demand amplification is the tendency of small fluctuations in demand at the retailer end of the supply chain to be amplified as they are communicated down the chain. A brief review of the literature on this phenomenon is presented, concentrating particularly on the causes propounded. A continuous-time differential equation model of a production-inventory system is then proposed. The application of a novel optimal control algorithm is applied in order to simulate the rational behaviour of inventory managers. This algorithm allows us to mimic the discontinuous cost structures implied by the advantages of batched production. By simulating the response of the system to small changes in demand, the relationship between batch size and the magnitude of demand amplification is investigated.

1. Introduction

Demand amplification was one of the first generic supply chain phenomena to be recognised (Forrester, 1961). Also known as the Bullwhip Effect, the Forrester Effect, or the Law of Industrial Dynamics, it is the proclivity of fluctuations in demand to be amplified as they are communicated away from the retailer.

* This work was supported by EPSRC under the grant number GR/M27432

In this paper we construct a dynamic model of a production-inventory system, whose novelty derives from the inclusion of a realistic cost structure implied by batched production. By employing a novel nonlinear optimal control algorithm we then attempt to simulate the rational behaviour of inventory managers. Through consideration of an idealised scenario we isolate and quantify the contribution of batched production issues to the demand amplification effect. We conclude that the relationship between batch size and demand amplification is not linear. It is, in fact, intimately correlated to the quotient (average demand rate)/(batch size).

The paper is structured as follows. In section 2 we give a brief review of the influential research on demand amplification which relates to our study. Then, in section 3, we discuss the feasibility and legitimacy of using differential equation techniques to model inventory systems and, more generally, supply chains, with particular emphasis on optimal control approaches. Section 4 introduces the optimal control algorithm. The remaining of the paper is devoted to the study of a particular scenario which is engineered to elucidate the connection we are interested in.

2. Demand Amplification

Demand amplification was first artificially reproduced using a continuous time simulation model by Forrester (1961). Typical amplification ratios which have been observed between two echelons are 2:1 (Towill 1992) and, between four echelons, 20:1 (Houlihan 1987). Here we have used the term echelon to describe any discernible generic activity in the supply chain. The Forrester Effect, which has also been called The Law of Industrial Dynamics (Burbidge 1984), has traditionally been attributable to a combination of factors. The following chain of events is a typical occurrence: Upswings in demand create a perceived shortage somewhere along the chain.

This may simply mean a fall in inventory below a target level. Lacking an overview of the entire supply chain, the company concerned then over orders to protect itself against further fluctuations. This increase in orders triggers further localised protection since it is misinterpreted as real extra orders. Houlihan (1987) produced a metaphorical interpretation for demand amplification by observing that the feedback loops inherent in these systems create a 'flywheel effect'. In this way relatively small ripples on level scheduling at the retail end of the supply chain can cause massive variations in demand at the raw materials end. Some factors that exacerbate demand amplification are listed below:

- Lead time delays.
- Unreliable delivery service compensated for by additional inventory investment.
- Uncertainty in the quality of information being passed between echelons.
- Poor demand forecasting. This may be unintentional or caused by tendentious 'forecasts' made by ambitious sales departments.
- Reducing product life cycles creating obsolete stock.
- Distinguishing periodic economic and seasonal demand swings from fundamental changes in consumer attitudes.

To counter demand amplification companies typically increase their buffer inventories in an attempt to smooth production rates. Unfortunately, if this is not done in a co-ordinated manner, every company in the chain can end up holding expensive levels of stock against the same contingency. Also, these extra levels of stock serve to cloud further the perception of any genuine demand fluctuations.

The excess material and manufacturing costs of sudden bursts of production, and the burden of capacity under-utilisation in concomitant idle periods motivates more detailed analysis of the specific drivers of demand amplification. To this end various authors have developed mathematical models of supply chains in an attempt to understand the 'pathology' of these drivers. The behavioural assumptions which drive these models represent different interpretations of the character of demand amplification. For instance in Sterman (1989) attributes a degree of irrationality to decision makers in the supply chain. In contrast, Lee et al. (1997) model the outcome of rational decisions among supply chain managers. In this paper they identify four drivers of demand amplification, one of which is order batching. This results from the advantage to retailers of economies of volume purchasing and transportation and leads to the investment in inventory stocks. Inevitably a postponement of ordering whilst inventories are depleted, and perhaps a toleration of stockouts follows as the price of securing such economies. Reconciling this trade-off is at the heart of inventory theory. In common with Lee et al. we shall assume rational and optimising behaviour of supply chain players. We shall approach the batch issue from its source, namely the discontinuous variation of production costs with volume.

3. Differential Equation Models

Modelling production-inventory systems and supply chains using differential equations holds great appeal for the control theorist. This is because many of the influential characteristics of the problem can be succinctly expressed in differential equation form. Then a vast array of tools and methodologies from control theory can be invoked to gain insight into the system dynamics. The rationale for this approach is that models of modest complexity, which are therefore amenable to analytical study, can provide an insight into the factors which are common to much larger 'live' systems. Differential equation models also have the advantage of being a conduit into the frequency domain, which offers a framework particularly suited to the study of systems in which oscillations are a salient attribute. There one can investigate which factors determine how various seasonal and other demand fluctuations may be amplified as they are passed along the chain.

Since differential equations produce 'smooth' outputs, they are not universally suited to the modelling of production-inventory system and supply chains. The system must be considered at an aggregate level, in which individual entities in the system (products) are not considered. Rather, they are aggregated into levels and flow rates, which vary smoothly with time. So these methods are unsuited to production processes in which each individual entity has an impact on the fundamental state of the system. For the same reasons differential equation approaches cannot solve lot sizing and job sequencing problems.

Towill and Del Vecchio (1994) regard the supply chain as a series of amplifiers, which are thus amenable to classical frequency domain techniques. They develop a three echelon simulation model of a supply chain in order to investigate the dynamics of demand amplification. The application of some heuristic design criteria is made and each individual echelon is tuned in isolation in order to mitigate this phenomenon.

Porter and Taylor (1972) dealt with the set of equations:

$$\begin{aligned} \frac{di}{dt} &= p_a(t) - d(t) \\ \frac{dp_a}{dt} &= \alpha\left(p_d(t) - p_a(t)\right) \end{aligned} \tag{1}$$

where $p_d(t)$ is the desired production rate, $p_a(t)$ is the actual production rate, $d(t)$ is the demand and $i(t)$ is the inventory level. By specifying a constant desired inventory level, these equations can be rearranged and the control inputs chosen to be the desired production rate and its derivative. Simple state feedback techniques are then developed to stabilise the system.

Bradshaw and Porter (1975) use similar modal control techniques in a slightly more complex environment in which advertising affects demand. This framework fails to take into account the cost implications of such control strategies. To do this we need to use the theory of optimal control.

Using standard optimal control theory over finite time intervals the various costs of the production strategy can be accounted for and traded-off against one another. The optimal control of such systems is tackled in the papers (Bensoussan and Proth 1982), (Abad 1985), (Lieber 1973) and the books (Bensoussan et al 1974) and (Arrow et al 1958). A common attribute of this work is that the faithful replication of real-world cost structures has been sacrificed in order to achieve tractable solutions. For example, both Bensoussan et al (1974) and Abad (1985) assume quadratic production and inventory holding costs, enabling them to invoke the relatively mature subject of linear-quadratic optimal control. Using these methods over a limited operating range does constitute a plausible approximation to linear cost structures. Further, away from the equilibrium, the accelerating cost penalties can legitimately be used to represent capacity constraints. However, large fluctuations in demand may require calculations over more extended cost ranges whose differing qualitative nature cannot so easily be defended. Arrow et al. (1958) use linear cost structures to investigate smoothed production plans when demand is known over a fixed planning horizon. The simplicity of linear costs engenders the achievement of significant analytical progress in this problem.

The paucity of papers treating more sophisticated cost structures bears witness to the difficulty of the problem. However, two papers (Bensoussan and Proth 1982), (Lieber 1973) have considered finite-time interval problems with simple convexity requirements on the costs. These may not always be fulfilled for particular processes and can be difficult to check, but seem to be the extent to which general conditions can be relaxed. Lieber uses Pontryagin's Maximum Principle to derive some planning horizon results. He requires the production cost function to be twice continuously differentiable, which may not be the case in reality as these costs tend to jump up with the number of production runs, not the number of goods produced. However, continuous approximations can be made to arbitrarily closely mimic these functions. A particular innovation by Bensousan and Proth (1982) is to take into account capacity constraints. They use backwards dynamic programming to calculate optimum production schedules. This method can be computationally burdensome and is only as good as the chosen algorithm. Neither of these papers gives examples which replicate the costs implied by batched production. A review of some of the literature mentioned here can be found in (Axsater 1985).

4. Optimal Control

The algorithm used in this paper is based on work first published in (Banks and Mhana 1992). They control systems of the form

$$\dot{x} = A(x)x + B(x)u, \tag{2}$$

using the infinite-time cost functional

$$J = \int_0^\infty \left(x^T Q x + u^T R u\right) dt . \tag{3}$$

Here x and u are the $n-$ and $m-$dimensional state and control vectors, respectively. A, B, Q and R are of the appropriate dimension and A and B are analytic. They show that under some bounded growth conditions on the derivatives of A and B, their algorithm produces a stabilizing control. The dynamic programming approach to the solution of this problem leads to a stabilising state feedback control of the form

$$u = -B(x)R^{-1}B^T(x)V_x(x), \tag{4}$$

where $V_x(x) \equiv \dfrac{\partial V}{\partial x}(x)$ satisfies the resulting form of the Hamilton-Jacobi-Bellman (HJB) partial differential equation (PDE)

$$V_x^T A(x)x - \frac{1}{2}V_x^T B(x)R^{-1}B^T(x)V_x + \frac{1}{2}x^T Qx = 0. \tag{5}$$

The problem of finding suitable solutions to equations like (5) has preoccupied mathematicians for years. The Banks and Mhana approach takes its inspiration from linear quadratic control, for which A and B are constant matrices in (2). Freezing the state in (5), they put $\dfrac{\partial V}{\partial x} = P(x)x$, where P is some matrix-valued function. Then (5) becomes.

$$x^T\left(P^T(x)A(x) + A^T(x)P(x) - P^T(x)B(x)R^{-1}B^T(x)P(x) + Q\right)x = 0. \tag{6}$$

Then, at any x, the algorithm consists of finding a positive-definite symmetric P such that

$$P(x)A(x) + A^T(x)P(x) - P(x)B(x)R^{-1}B^T(x)P(x) + Q = 0. \tag{7}$$

The condition for existence of such a solution is that the frozen linear system $(A(x), B(x), Q)$ be stabilizable and observable for all x on the stable trajectory. This condition is difficult to prove *a priori,* but, for polynomial functions, its failure will only be ephemeral. The utility of this algorithm stems from its simplicity and the resulting ease of application. At any x along a trajectory one is only required to solve the algebraic Riccati equation. That the solution of the HJB equation is obviated is its clearest benefit.

We must point out that after Banks and Mhana published their work there appeared a paper (Gong and Thompson 1995) questioning the reasoning behind one of the proofs in the multivariable case. These matters were resolved in (McCaffrey and Banks 1998), where the algorithm was shown to be asymptotically optimal (that is, converges to the optimal control as $x \to 0$) in this case.

Both these last two references extend the problem to consider analytic state dependent cost matrices (Q and R in (3)). It is this innovation that allows us to model batched production costs and capacity constraints. In practice we have allowed the failure of continuity at isolated points by considering piecewise-linear-affine functions to achieve the most realistic cost structure. By choosing the algorithm's integration step-length small enough we have found this not to cause any problems. In any case it is possible to approximate arbitrary closely these functions by smooth polynomials or trigonometric functions. In the paper (Harrison and Banks 1998) the authors use this algorithm to control a vehicle suspension system. They too use discontinuous cost matrices in order to better approximate the engineering objectives of the system. The approximation of hard constraints by the appropriate modification of the cost functional is discussed in more detail in (Russell 1965).

Our model of an echelon in the supply chain has an external forcing function (the demand) as well as a feedback input. It is well known that, for constant inputs such systems reduce to the familiar regulator problem (with no external input). For the purposes of this study we shall only consider constant inputs by assuming that fluctuations in demand are initially satisfied by a safety stock of inventory.

5. Modelling Supply Chains

Consider a production-inventory system modelled by the following set of equations:

$$\frac{di}{dt} = P_a(t) - d(t), \qquad \frac{dP_a}{dt} = \alpha[P_d(t) - P_a(t)], \tag{8}$$

where $d(t)$ is the demand received from the retailer and $i(t), P_d(t), P_a(t)$ denote the inventory level, the desired production rate (the control variable) and the actual production rate of, respectively. The latter two are related by a simple exponential time delay with parameter α. The simulation of rational and optimising decisions by the inventory manager is achieved by the application of the optimal control algorithm. The cost functional is of the form

$$J = \frac{1}{2}\int_0^\infty \left(q_1(i) \times (i)^2 + q_2(P_a) \times (P_a)^2 + r \times (P_d)^2\right) dt, \tag{9}$$

for $q_1, q_2 \geq 0, r > 0$. The components of the integrand are contrived to look like

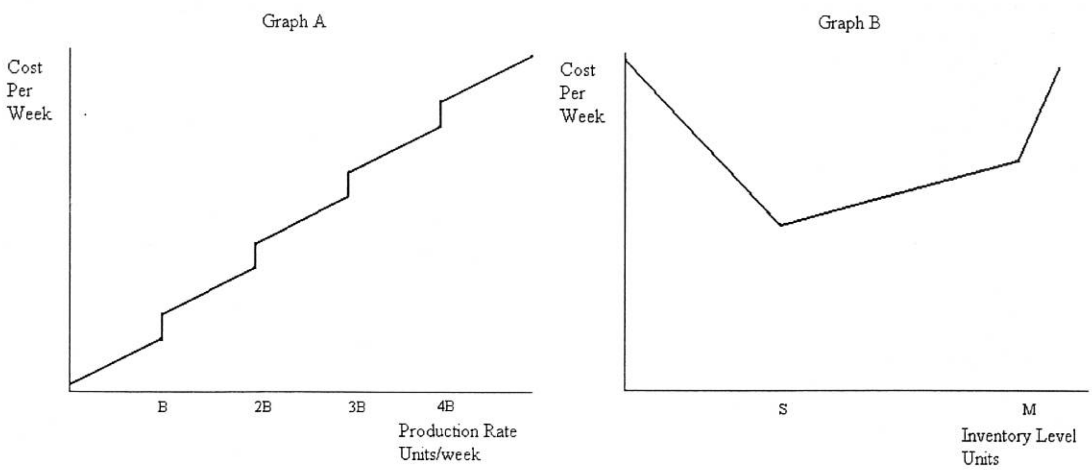

Figure 1. Production and Inventory Cost Structures

Figure 1 (Graph A) shows how the cost per unit time of production varies with the production rate. This comprises a fixed batch setup cost of c_1 and a marginal unit cost of c_2 per unit rate of production (batch size $= B$). Mathematically it is possible for P_a to become negative. However for physically realistic systems this remains a pathological possibility.

Inventory cost structures (shown in Graph B) tend to be rather notional in character since the calculation of the opportunity cost of holding inventory is based on an element of supposition (see Lee and Billington 1992). S is the desired safety stock level. In addition to the opportunity cost of the capital and a quantification of the risk of obsolescence, this incorporates the fixed cost of the warehousing facility. Above this level storage costs increase linearly with their level (at a rate c_5 per unit) up to the maximum capacity M, after which the cost increases precipitously to deter incursions into this 'forbidden zone'. Below S we are penalising the risk of stockouts at a rate of $-c_4$ per unit.

It is possible to consider the possibility of stockouts in this model, when $i < 0$. However, for simplicity of exposition we assume that S is chosen large enough to avoid unsatisfied demand.

In order to avoid both functions q_1 and q_2 'blowing up' near zero we have redefined them to be constant around this point. Choosing this region to suitably small we have found that this does not affect the optimal policy.

The parameter r is chosen to have negligible impact on the overall cost structure.

6. The Effect of Batch Size on Demand Amplification

In this section we study the effect of batch size on the production rate following a perturbation in demand. In order to distinguish most clearly this effect we postulate a scenario in which the firm is working to a level schedule, i.e. demand is assumed to fluctuate about a constant level in the medium term ($d(t) = d$). We are interested in the consequences of a sudden small increase in demand over this level. This sort of instantaneous change is almost discrete in character. Hence the most realistic way to model it is by a drop in inventory below the safety stock level.

In order to contrive the realistic behaviour of the system we append a notional fixed cost to the inventory cost structure. This ensures that at the 'average' demand rate, d and with the inventory level at S, the production and inventory costs are equal. So the production rate remains constant. In reality this notional extra cost is manifest in the calculation of the level S (which is at the discretion of the production manager) and depends on the magnitude and variability of demand.

These considerations are designed to yield a system in 'equilibrium' over the medium term. They therefore aid the discrimination between the contribution of batch size to demand amplification and other known causal factors.

The following parameter values characterise the system:

$$c_1 = £250, c_2 = £0.50 \text{ per unit per week}, c_4 = -£400 \text{ per unit per week}, \alpha = 2,$$
$$c_5 = £0.70 \text{ per unit per week}, S = 1000 \text{ units}, r = 1/1000, d = 1800 \text{ units per week}.$$

The capacity constraint M is taken to be large enough so as not to have an influence on the dynamics. The batch size B is varied between the values $B = 350....1400$. The fixed inventory storage cost varies with B. The initial conditions of the system are $i(0) = 0, P_a(0) = 1800$. Hence an instantaneous demand has depleted the safety stock level to zero.

Figure 2A shows the resulting production rate when $B = 800$. The corresponding inventory levels are shown in figure 2B.

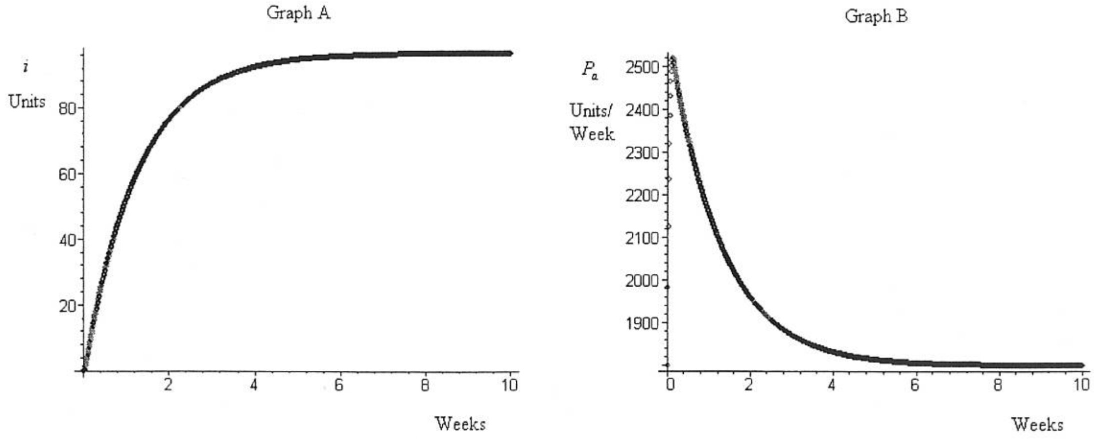

Figure 2. Production and Inventory over time for example

By running similar simulations for varying batch sizes and noting the maximum production rate of each we are able to plot a measure of demand amplification against batch size (see figure 3). This is a reasonable statement since the demand placed upon suppliers of our firm can reasonably be assumed to be proportional to the rate of production (See Towill 1992).

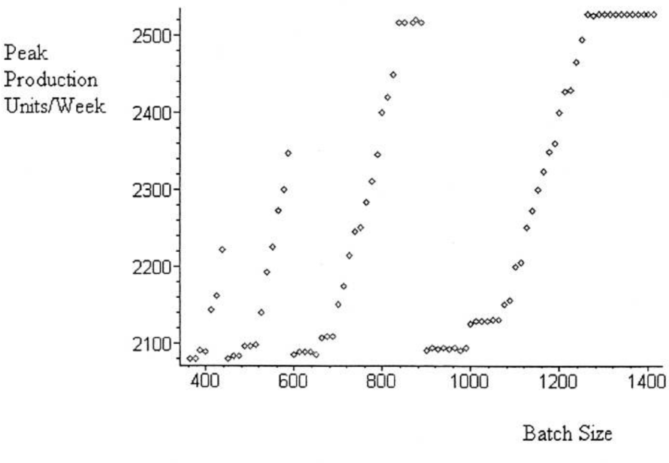

Figure 3. Measure of Demand Amplification for various batch sizes

A number of conclusions can be drawn from this example. Firstly, the relationship between batch size and demand amplification is not linear. In general we cannot say that smaller batches lead to lower amplification. There are certain intervals in the range of batch sizes in which demand amplification is considerably suppressed. These occur in ranges following batch sizes which are divisors of the average demand rate. In the example these are $450 < B < 500, 600 < B < 700$ and $900 < B < 1000$. Hence the level of demand is an influential factor in this study. Above the largest divisor of d (900) amplification increases up to a certain point and then levels off.

7. Conclusions

In this paper we have quantified the variation of demand amplification with batch size. To do this we set up a simple production system satisfying a demand which, in the medium term, averaged to a constant. Then we applied a novel optimal control technique to simulate the rational behaviour of inventory managers. The algorithm used found the optimal production rate over time when the firm was confronted with fluctuations in inventory safety stocks. The demand transmitted to the next lower echelon in the supply chain was assumed to be proportional to our firm's production rate. Hence a credible measure of demand amplification was found to be the maximum production rate following a small instantaneous drop in inventory. By plotting this rate for varying batch sizes we determined a relative measure of the connection between batch size and demand amplification.

This relationship was found to be intimately connected to the divisors of the demand rate. At batch sizes that were slightly larger than these divisors amplification was suppressed. We venture that this algorithm might prove useful to inventory managers in understanding the mechanism of demand amplification and its suppression.

It should be noted that since we were attempting to isolate and quantify the influence of batch size on demand amplification, we created an idealised scenario in which other causal factors were artificially subordinated. For example, the time delay was chosen to be relatively small. Also the demand fluctuations were assumed to be around a level demand (in the medium term) and initially satisfied be safety stocks. Perhaps more importantly, we attributed rational and optimising behaviour to inventory managers. The relaxation of all these assumptions should cause amplification to be increased.

Abad, P.L., Two-level algorithm for decentralized control of a serially connected dynamic system, *Int. J. Systems Sci.*, **16**, 5, (1985), 619-624.

Arrow, K.J., Karlin, S., Scarf, H., Studies in the mathematical theory of inventory and production, (Stanford: SUP), 1958.

Axsater, S., Control theory concepts in production and inventory control, *Int. J. Systems Sci.*, **16**, 2, (1985), 161-169.

Banks, S.P., Mhana, K.J., Optimal control and stabilization for nonlinear systems, *IMA J. Math Con. & Inf.*, **9**, (1992), 179-196.

Bensoussan, A., Hurst, E.G. Naslund, B., *Management applications of modern control theory*, (Oxford: North-Holland), 1974.

Bensoussan, A., Proth, J.M., Inventory planning in a deterministic environment continuous time model with concave costs, *Euro. Inst. Adv. Studies in Man.*, working paper, April 1982.

Bradshaw, A., Porter, B., Synthesis of control policies for production-inventory tracking system, *Int. J. Systems Sci.*, **6**, 3, (1975), 225-232.

Burbidge, J.L., Automative production control with a simulation capability, Paper presented at IFIP conference, Copenhagen, 1984.

Forrester, J.W., *Industrial dynamics*, (Cambridge Mass: MIT press), 1961.

Gong, C., Thompson, S., A comment on 'stabilization and optimal control for nonlinear systems', *IMA J. Math Con. & Inf.*, **12**, (1995), 395-398.

Harrison, R.F., Banks, S.P., Towards real-time, nonlinear quadratic optimal regulation, *Proc. IFAC Algorithms and Architectures for real-time control, AARC'98, Cancun, Mexico*, 417-422, 1998.

Houlihan, J.B., International supply chain management, *Int. J. Pys. Dist. Mat. Man.*, **17**, 2, (1987), 51-66.

Lieber, Z., An extension to Modigliani and Hohn's planning horizons results, *Management Science*, **20**, 3, (1973), 319-330.

Lee, H.L., Billington, C., Managing the supply chain inventory: pitfalls and opportunities, *Sloan Man. Rev.*, (spring 1992), 65-73.

Lee, H.L., Padmanabhan, V., Whang, S., Information distortion in a supply chain: the bullwhip effect, *Management Science*, **43**, 4, (1997), 546-558.

McCaffrey, D., Banks, S.P., Lagrangian Manifolds and asymptotically optimal stabilizing feedback, Research Report No. 722, Dept Automatic Control & Systems Engineering, University of Sheffield, 1998.

Porter, B., Taylor, F., Modal control of production-inventory systems, *Int. J. Systems Sci.*, **3**, 3, (1972), 325-331.

Russell, D.L., Penalty functions and bounded phase coordinate control, *J. SIAM*, **2**, 3, (1965), 409-422.

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

Towill, D.R., Supply chain dynamics-the change engineering challenge of the mid 1990s, *Proc. Instn. Mech. Engrs.*, **206**, (1992), 233-245.

Towill, D.R., Del Vecchio, A., The application of filter theory to the study of supply chain dynamics, *Prod. Plan. & Con.*, **5**, 1, (1994), 82-96.