



Incorporating demand uncertainty and forecast error in supply chain planning models[☆]

R Fildes* and B Kingsman

Lancaster University, Lancaster, UK

This paper develops a framework for examining the effect of demand uncertainty and forecast error on unit costs and customer service levels in the supply chain, including Material Requirements Planning (MRP) type manufacturing systems. The aim is to overcome the methodological limitations and confusion that has arisen in much earlier research. To illustrate the issues, the problem of estimating the value of improving forecasting accuracy for a manufacturer was simulated. The topic is of practical importance because manufacturers spend large sums of money in purchasing and staffing forecasting support systems to achieve more accurate forecasts. In order to estimate the value a two-level MRP system with lot sizing where the product is manufactured for stock was simulated. Final product demand was generated by two commonly occurring stochastic processes and with different variances. Different levels of forecasting error were then introduced to arrive at corresponding values for improving forecasting accuracy. The quantitative estimates of improved accuracy were found to depend on both the demand generating process and the forecasting method. Within this more complete framework, the substantive results confirm earlier research that the best lot sizing rules for the deterministic situation are the worst whenever there is uncertainty in demand. However, size matters, both in the demand uncertainty and forecasting errors. The quantitative differences depend on service level and also the form of demand uncertainty. Unit costs for a given service level increase exponentially as the uncertainty in the demand data increases. The paper also estimates the effects of mis-specification of different sizes of forecast error in addition to demand uncertainty. In those manufacturing problems with high demand uncertainty and high forecast error, improved forecast accuracy should lead to substantial percentage improvements in unit costs. Methodologically, the results demonstrate the need to simulate demand uncertainty and the forecasting process separately.

Journal of the Operational Research Society (2011) 62, 483–500. doi:10.1057/jors.2010.40

Published online 16 June 2010

Keywords: lot sizing; forecasting; service levels; demand uncertainty; value of forecasting; supply chain

1. Introduction

Interest by researchers and practitioners in understanding the consequences of uncertainty and forecast error for manufacturing and supply chain activities remains strong after over 25 years of activity (Dolgui and Prodhon, 2007). The topic is important to practitioners because companies throughout the supply chain spend large sums of money on purchasing, implementing and operating forecasting systems and they need to justify such expenditure and get the most out of these systems. Academics have focussed throughout the period on the effects of demand uncertainty on manufacturing systems

(eg De Bodt and Van Wassenhove, 1983a; Jeunet, 2006) and the Bullwhip effect in the supply chain (Forrester, 1961; Lee *et al.*, 1997), sometimes incorporating aspects of forecasting error (Chen *et al.*, 2000a, 2000b). More recently, both groups have become interested in collaborative planning and forecasting where information is shared between companies to co-ordinate production and distribution. Although some researchers have proposed analytical models to test various hypotheses in these problem areas, they are inevitably forced to make simplifying assumptions to ensure mathematical tractability and therefore cannot deliver estimates of key parameters, such as the value of collaboration or forecasting accuracy. In this paper we propose a framework for simulating the effects of forecast error, focussing on a key managerial issue: the value of improved forecasting accuracy in a manufacturing system.

In order to estimate the value of forecasting, we have simulated an MRP system for co-ordinating production and operations planning. Such systems typically assume a deterministic environment with the emphasis on co-ordinating production at all levels and the purchasing of components and materials

[☆]This paper represents part of a long-standing research collaboration between Robert Fildes and Brian Kingsman who sadly died before it could be completed. A preliminary version of aspects of this research has been available since 1997 discussing some of the same issues. Valuable comment on this new material has been received from July Jeunet and Alan Mercer but any remaining confusion remains Robert Fildes' responsibility.

*Correspondence: R Fildes, Department of Management Science, Lancaster University, Lancaster LA1 4YX, UK.

E-mail: R.Fildes@Lancaster.ac.uk

so as to meet the planned Master Production Schedule (MPS). The MPS provides the linkage between the demands of the outside customers and the detailed planning and scheduling of internal production supply facilities. The MPS, once determined, is passed onto the material requirements planning system. Uncertainty can affect the performance of the MRP system in a variety of ways. Random variations can occur in internal manufacturing lead times, in purchasing lead times or result from stochastic yield variations at production stages. As these activities mainly fall within the span of control of the company management, they can be managed to reduce significantly the amount and impact of uncertainty. External demand from customers is, however, normally outside the company's control, except for gross influences like promotions, and is stochastically varying over time in most practical situations. Thus the MPS has to be developed on the basis of forecasts of demand.

The scheduled production will either be insufficient to meet the realised actual demand, thus producing shortages, or exceed the realised demand resulting in unanticipated stocks. As several authors have pointed out, for example Vargas and Dear (1991), the effect of demand uncertainty on manufacturing systems that rely on deterministic lot-sizing techniques, such as MRP applications, is an important problem. Although all researchers seem to agree with the early conclusions of De Bodt and Van Wassenhove (1983a) and others that demand uncertainty dramatically increases costs, the simple questions, of how those costs increase with forecasting error, their relationship to demand uncertainty and whether the form of demand pattern and the forecasting model used are important, remain unresolved.

In order to investigate this problem area researchers must make a number of choices: (i) on the form of the MRP system, including the number of levels and the interdependence within the production process, capacity constraints, production/delivery lead times, and back-ordering rules; (ii) the lot sizing rules (LSRs) to be considered and their application at differing levels of the MRP system; (iii) the costing of the various activities, finished inventory, work in progress, ordering; and (iv) service level and buffering strategy. In this paper we explore a number of important, previously neglected, issues, including the form of the demand and forecast functions, the levels of forecast error to be included in the experimental design and the effects of forecast mis-specification.

Researchers have tackled the problem of uncertainty and forecast error in MRP systems in two distinct ways: firstly, through case studies with changes in forecast errors arising from the use of different forecasting methods (eg, Gardner, 1990; Lee *et al.*, 1993; Enns, 2002) and, secondly, through developing a simulation of the whole MRP system, including the demand distribution and forecast errors (eg, Jeunet, 2006). This paper adopts the second approach with the aim of developing a satisfactory framework for quantifying the effects of forecast error and mis-specification.

As an illustration of the proposed methodology we examine a single product two-stage production system using various lot sizing rules applied within the MRP logic. The aim of the model is to estimate the value of improved forecasting. The methodology we describe, however, extends straightforwardly to more complex production systems with manufacturing uncertainty, collaborative planning (Aviv, 2001), interdependent demand at the final product or component level and capacity constraints (Xie *et al.*, 2004). This paper is in six further sections. We next consider demand uncertainty and the corresponding optimal forecast. Section 3 describes the simulated manufacturing system including costs and LSRs while Section 4 describes the simulation experiment. The substantive results of the simulation are presented in Section 5 for differing levels of demand uncertainty. Section 6 considers the effects of different forms of forecast error and forecasting model mis-specification (in addition to the demand uncertainty). The paper concludes with a discussion of the results, both substantive and methodological.

2. Demand uncertainty and forecast errors in supply chain systems

Many studies have evaluated the performance of deterministic lot-sizing rules in manufacturing systems. The effects of uncertainty on manufacturing systems, particularly final demand uncertainty, have also generated considerable interest, with Wemmerlöv (1989) offering an early summary. It is endemic to most organisations and can be reduced but not eliminated by accurate forecasting. Failure to include uncertainty in an analysis severely limits its practical implications. The earlier research suffered from limitations that include a lack of clarity in distinguishing between demand uncertainty and forecast error (De Bodt and Van Wassenhove, 1983a; Lee and Adam, 1986) and the trade off between cost and service level, which itself depends on the cost structures assumed in the problem (Vargas and Dear, 1991).

Although much of the research analysed a simple production system (eg De Bodt and Van Wassenhove, 1983b; Lee and Adam, 1986; Ritzman and King, 1993; Chen *et al.*, 2000a) a few researchers have considered manufacturing systems closer to the complexity observed in practice (eg Vargas and Dear, 1991; Jeunet, 2006). The problem with more complex studies is that the quantitative results are likely to be specific to the distinctive features of the chosen production system. Although not discounting the need to examine complex systems we will follow the approach adopted by Wemmerlöv and Whybark (1984) and Wemmerlöv (1989). These studies examined a single product production system in a rolling horizon environment and included a range of LSRs. The conclusion of Wemmerlov's research gave support to De Bodt and Van Wassenhove (1983a)—that an inventory system facing uncertain demand is fundamentally different to one with a known demand and that results from the latter case will not apply in most practical situations.

Using Wemerlov's framework permits us to generalise his experimental setting. In an attempt to overcome the major problem of service level associated with most of the previous work, Wemmerlöv (1989) partially controlled for service level by using a buffer to ensure that it was 100%. In the research described here we go further including service level as a continuous variable. This permits us to evaluate the effects of realistic improvements in forecasting accuracy on differing levels of service and unit costs.

The inadequate incorporation of service level has led to further confusion in past research: that of failing to distinguish between the (unknown) bias in the forecast and a buffering strategy of forecast inflation that aims to moderate the effects of forecast errors by increasing the forecast (Lee and Adam, 1986), thereby providing an inventory buffer. Such a buffering strategy has long been criticised as inadequate in manufacturing systems when compared to the alternative of using safety stocks (Vargas and Dear, 1991). Here we use safety stocks in the final product to ensure a variety of service levels.

2.1. Demand models used in supply chain research

Early authors when examining the effects of uncertainty on manufacturing systems considered only limited models of the demand generation process and forecast function. De Bodt and Van Wassenhove (1983a, 1983b) compared various LSRs when demand was generated by $D_t = \mu_d + \varepsilon_t$ where $\{\varepsilon_t\}$ is assumed independent and identically distributed, $N(0, \sigma^2)$ and μ_d is constant. The demand forecasts were derived using exponential smoothing. They concluded that with any level of demand uncertainty greater than zero total costs were increased by about 10%; the impact was highest when ordering costs were low compared to inventory costs (and the time between orders was therefore short). Also, the choice of LSRs was unimportant. Mean demand was set at 200 and σ varied from 0 to 50 giving corresponding coefficients of variation from 0 to 0.25. The MRP system was such that current demand was known and could be immediately supplied. Unrealistically, there were consequently no stock-outs. They also fail to match their generation process for demand (random about a constant level) to their forecasting method (exponential smoothing). This is a common mistake, which compounds forecasting error with mis-specification error (Newbold and Bos, 1989) and makes interpretation of the results difficult since they constitute a mixture of the mis-specifications and the various experimental effects.

Lee and Adam (1986) and Lee *et al* (1987) examined more complex MRP systems. Demand was generated by $D_t = \mu_d$, a constant while the forecasts were seen as randomly distributed around μ_d , as $N(\delta, \sigma^2)$ where δ is the bias in the forecast. A wide range for the coefficient of variation was considered. Despite the range, the error standard deviation had little effect on performance while the bias was evaluated as much more important. The conclusion they drew that the bias was the more important factor again arose from the limited cost

parameters considered. More recently, research has focussed on bullwhip effects and supply chain coordination (see eg, Chen *et al*, 2000a, b; Xu *et al*, 2001; Zhao *et al*, 2001, 2002). Despite these studies carefully specifying both the data generating process (DGP) and the forecasting model, one of their fundamental and unanalysed assumptions is that the forecasting model used by the manufacturer is mis-specified. Lee *et al* (2000) and Aviv (2001, 2002), in contrast, match the two assuming a stationary autoregressive process, an issue we return to in the next section.

A summary of selected earlier research to 2009 is shown in Table 1 spanning the earliest studies that proved influential in specifying the problem area to a range of recent contributions across a wide range of manufacturing systems. This clearly demonstrates the need for a methodology that incorporates the factors we have identified in the literature survey into the experimental design: service level, cost structure, the form of the demand generation process and the forecasting process. Although individual papers have included some of these experimental factors it still remains true, 25 years after the original research, that the failure to establish an explicit experimental design leads to confounding of these factors and possibly mis-leading conclusions. This lack of generality is particularly damaging when quantitative estimates of effects' sizes are sought, which, as we argued earlier, are a managerial necessity if a new forecasting system is being bought or a Collaborative Planning, Forecasting and Replenishment arrangement is being evaluated.

2.2. Independent demand and corresponding forecasting methods

Although a plethora of different forecasting methods have been proposed as suitable for forecasting independent demand, recent large scale empirical studies have concluded that a limited range of simple methods perform as well as their more complex alternatives (Fildes *et al*, 1998; Makridakis and Hibon, 2000). Particularly for monthly or weekly data on item demand, autoregressive statistical models should adequately characterise most of the situations encountered in practice. Two simplifications from the typical series observed in practice have been made here: extreme (or outlying) observations have been neglected and the data are assumed to be both non-seasonal and with no trend. For short lead times the former assumption is the more constraining. The effects of these assumptions will require further research. We therefore propose to analyse the following demand model, where D_t is the demand in period t and e_t is random white noise:

$$D_t = \delta + \rho D_{t-1} + e_t - \theta e_{t-1} \quad |\rho| \leq 1, |\theta| < 1 \quad (1)$$

This model, which is a realistic generalisation of earlier work, includes a wide range of data types that are commonly observed in practice as the forecasting literature has established. For $\vartheta = \rho = 0$ the model is just the 'constant mean

Table 1 Characteristics of selected earlier research; 1984–2009

<i>Characteristic</i>	<i>Ideal (Current study)</i>	<i>Jeunet (2006)</i>	<i>Aviv (2001, 2002)</i>	<i>Enns (2002)</i>	<i>Chen (2000a, b)</i>	<i>Lee and Adam (1986)</i>	<i>Wemmerlöv and Whybark (1984), Wemmerlöv (1989)</i>
Manufacturing system/ supply chain	Experimental factor (Simple)	Simple/ Complex	Two stage supply chain	Complex MRP for 2 products	Two stage supply chain	Simple/ Complex MRP	Simple single level
Focus	Applicable to all types of problem (Yes)	Estimated effects of forecast error	Tests of benefits of collaborative forecasting	Forecast error and stock policies	Bullwhip effect and benefits of information sharing	Tests of the effects of forecast error	Tests of Effects of forecast error
Service level	Experimental factor (Yes)	Fixed at 100%	Endogenous	System output measure	n.a.	Included only through shortage costs	Fixed at 100%
Cost Structure	Experimental factor (Yes)	Time between orders is an experimental factor	Fixed	Fixed	n.a.	Fixed	Time between orders is an experimental factor
Demand process	Experimental factor (Yes)	Deterministic	Autoregressive + market signal	Deterministic with seasonal mean	Autoregressive (linear trend)	Constant	Random
Forecasting process	Optimal, non-optimal & mis-specified (Yes)	Random around DGP	Optimal	Mis-specified Random around DGP	Mis-specified moving average (or exponential smoothing)	Random around constant	Random around demand
Replenishment policy	Experimental factor (Yes)	Various deterministic lot sizing rules	Order-up-to	Fixed order quantity	Order-up-to	Various deterministic lot sizing rules	Various deterministic lot sizing rules
Approach	Either analytic or simulation (Simulation)	Simulation	Analytic + numeric	Simulation	Analytic	Simulation	Simulation

demand model' (contrast the Lee and Adam's formulation described earlier) while for $\rho = 1$ and $\theta < 1$ the model is the ARIMA (0,1,1) model for which exponential smoothing is the optimal (minimum mean squared error) forecasting method. For $\rho < 1$ and $\theta = 0$ the model is a stationary AR(1) process and δ is determined by the mean of the process. The above model is the basis of the most common forecasting methods used in production systems and here we consider two cases: the AR(1| $\rho = 0$), a Random Noise series, and AR(1| $\rho = 0.9$). The latter model was chosen to simulate series close to non-stationarity and therefore close to a series where exponential smoothing is near-optimal.¹

The optimal (in the minimum mean squared error sense) one-step ahead forecast of demand for period t , ${}_{opt}F_{t-1,t}$ made in period $t - 1$ after that demand is known, for the above data generation process is given by taking expectations of Equation (1):

$${}_{opt}F_{t-1,t} = \delta + \rho D_{t-1} - \theta e_{t-1} \quad (2)$$

The error term, e_{t-1} , is the forecast error observed in the previous period, defined as the difference between the realised demand at $t - 1$ and the previous forecast, and at time t this is known. In general, of course, such an optimal forecast would be unavailable to the scheduler, not least because the parameters would have to be estimated from the data. In this study we assume that the parameters are known exactly; the only difference between the optimal forecast and the data generation process arises because of the random error, e_t .

The distribution assumed for e_t includes a range of values for the noise variance but the mean is assumed to be zero. The reason for this choice is that a constant bias model (such as that used by Lee and Adams, 1986) would be forecastable over time. The noise distribution is assumed normal (although demand and the forecasts of demand are constrained to be positive). The general form for the forecast error variance for an AR(1) model for k steps ahead is:

$$\text{var } e_t(k) = \frac{(1 - \rho^{2k})}{1 - \rho^2} \sigma^2 \quad (3)$$

For the Random Noise model, with $\rho = \theta = 0$, the forecast error variance is constant for all steps ahead. For the AR($\rho = 0.9$) model the standard deviation increases with k .

The complete specification of this data generation process depends on the values chosen for the noise variance. The values used in the experiment represent a plausible range of values for fast moving items according to surveys of forecasting accuracy reported in the literature (Fildes and Beard, 1992; Fildes and Goodwin, 2007; Fildes *et al.*, 2009). For one-step ahead (3 months ahead or less) forecasts, these studies suggest a mean absolute percentage error of at least 20%,

often higher. For a normal distribution this roughly translates into a coefficient of variation of 0.25. In the simulation experiment we will therefore use values of the coefficient of variation ranging from 0 to 0.4. For e_t assumed normal the range of values of the noise standard deviation included in the simulation experiment are therefore: 0, 50, 100, 200, 400 (around a mean of 1000) with demand constrained to be non-negative.

2.3. Forecasting error and mis-specification

Forecast error is usually taken as the difference between the actual and the forecast value. However, this error combines the randomness in the process generating the demands and the errors arising from not using the optimal forecast; most earlier researchers in this area have conflated the two. It is only the latter that represents the potential value of improving forecasting accuracy and the value of improved forecasting derives from the difference between the optimal forecasting method for a given DGP and whatever sub-optimal method is used. The non-optimal forecasting system can now be defined as:

$$F_{t-1,t+k} = {}_{opt}F_{t-1,t+k} + v_t \quad (4)$$

for each of the DGPs so that the actual forecast equals the optimal forecast plus Random Noise (assuming the same error is made across lead times—other assumptions could be justified). It follows that the overall forecast error is:

$$D_{t+k} - F_{t-1,t+k} = e_{t+k} - v_t \quad (5)$$

This distinction between randomness in the data generation process and the consequential forecasting errors permits the introduction of a wide range of forecasting error characteristics, including bias and forecast model specification error. Forecast mis-specification is a common feature of many relevant studies (eg Chen *et al.*, 2000a, b; Lee and Adam 1986).

In this paper we examine the case where v_t has error standard deviation $\kappa\sigma$ where κ is chosen so that the overall error standard deviation, $\sigma\sqrt{(1 + \kappa^2)}$, takes values of 110, 120 and 150 of the minimum attainable, given by σ . The results from introducing forecast error should look similar for different levels of randomness in the data generation process alone despite the two sources of error being independently simulated. (This will not in general be true in the common case where e_t and v_t are correlated due to a common cause of both demand uncertainty and forecast error such as a promotional campaign.)

Thus, there are two sources of uncertainty in the demand prediction process. One, the forecasting error, can be reduced by getting closer to the optimal forecasting method for the underlying demand generation process (DGP). Clearly, as there will be inevitable errors in identifying and estimating the parameters of the DGP, it will never be possible to eliminate such forecasting errors altogether. The second source, the process error, is the random variation in the DGP itself. This can be reduced only by attempting to manage the demand

¹ A referee queries why an ARIMA (0,1,1) model has not been considered. The focus of this already long paper is methodological and the proposals carry over to this important case; however, further research into more general demand processes is desirable.

process, for example by attempting to change customers' behaviour or through collaborative forecasting. The use of specific ordering days in a week or month for different groups of customers in inventory/distribution systems offers an illustration. In order to guide management effort towards process improvements it is important to distinguish as much as possible between the impact of these two sources of uncertainty on unit costs and service level. In the simulation study that follows we examine the consequences for costs and service level of different levels of demand uncertainty, in the first instance, supplemented by forecast error in the second.

3. The production and lot sizing environment

The production system simulated here is composed of two stages, a raw material (or component) having a purchasing and delivery lead time and a manufacturing process (or assembly) to make the final product from the material (or component), requiring one or more periods to be completed. The major areas of interest in the simulation experiment are the costs of producing the final product and the customer service level achieved, together with the effect of demand forecast errors on the MPS. The research presented here follows most earlier work in assuming that the only stochastic variation is in the demand for the final products. It is assumed that internal manufacturing lead times and purchasing lead times are fixed and also that internal yields are always 100%. The components are pushed up to the final level as scheduled. They are not delayed because of the fact the demand for the final product may change. All variation is absorbed into the stock of finished items.

3.1. Planning horizon and action horizon

In uncertain environments, MRP is usually developed using a rolling schedule. When using a rolling schedule, the production schedule is re-planned at regular intervals. Decisions are made taking into account the forecast demands over some future fixed planning horizon. In each successive period, lot-sizing decisions are made to minimise costs over the planning horizon ahead from that period. Then orders for raw material are placed and production occurs. An ordering decision by the manufacturer now leads to the scheduled arrival of that order some lead time into the future. Here it is assumed that once orders have been placed, they cannot be cancelled or altered in any way. Thus a portion of the production schedule determined for the planning horizon is frozen each planning cycle. This will include all those periods with the scheduled arrival of an outstanding order placed in previous periods. This frozen portion will be at least the lead time, L , to produce the item through all of the possible stages of purchasing and manufacturing. It may also include a further action horizon, a length of time, of say A periods, over which any recommended orders are made firm orders, which cannot then be changed in later periods. Given the assumptions of constant processing and lead times,

only the total length of the frozen period matters, and not how the frozen period is decomposed into the maximum manufacturing lead time and the action horizon.

The planner can only make changes to ordering decisions for that part of the planning horizon beyond the frozen period, the free planning horizon. The length of the free horizon dictates the number of periods over which demands can be batched into different lot sizes. The longer it is the greater the possibility of reducing costs by allowing bigger orders where advantageous. The total planning horizon is the frozen plus free periods. The simulation model allows different values for both the frozen and free horizons to be examined. The experiments covered values of 2 (the minimum possible for a two state production process), 5 and also 7 periods for the frozen period to gain some understanding of any sensitivity in the results. The free horizons considered were 12, 24 and 36, again to understand whether the results proved sensitive to this choice.

3.2. Costs

The costs included in this experiment consist of the cost of placing orders and the cost of holding stock. There are two major measures of performance: (i) the average ordering and stockholding costs per unit of demand, and (ii) the service level achieved, defined as the proportion of demand satisfied immediately either from the opening stock or the current period's output.

As most of the deterministic LSRs aggregate the net requirements over future time periods to give the lot sizes, the expected time between successive orders is an appropriate method of ranking alternative sets of cost parameters, the ordering cost and (end of period) stockholding cost. This was determined from the order cycle given for the classical Economic Order Quantity (EOQ), based on the average demand per period implied by the demand generation processes used. It is often known as the Time Between Orders (TBO) measure. The stockholding cost was set at 1 and the expected average period demand at 1000. Ordering cost values were chosen to give TBO values, measured in periods, in the middle of the ranges 1–2, 2–3, 3–4 and 4–5. These values imply the order would be expected to cover 2, 3, 4 and 5 periods' demands respectively. The corresponding values for the ordering cost will then be 1750, 3500, 7000, and 11500.

3.3. Service level and safety stocks

A service level performance measure has been used rather than including stockout costs as part of the total cost, since the latter are rarely known in practice. We adopted the approach of letting the service level be a factor in the experiments. This allowed us to examine the trade-offs between costs and service level. Clearly, under the same cost parameters and forecast error conditions, different LSRs result in different performance costs and different service levels. To render the

results comparable, the models were run with different levels of target end of period final product stocks. If this target level is a positive quantity then this is equivalent to holding safety stock. Lower service levels are only possible by deliberately forcing the system on average to have stockouts. This can be achieved by allowing backorders or lost sales up to some pre-set level, that is, by making the safety margin negative.

3.4. Lot sizing model for backorders and lost sales cases

As discussed above, the existence of an action horizon, A , means that at any time, t , recommended orders to be placed within the next A periods are made up of firm orders which cannot be changed at a later time. Thus at time t , orders already placed for periods $t+1$ to $t+A-1$ are outstanding orders that were placed at an earlier time and cannot be altered in any way. These orders will arrive a lead time L after they were placed, in periods $t+1+L$ to $t+A-1+L$. Let us denote the frozen interval as discussed above, the sum of the action horizon and the lead time in periods, as $F=A-1+L$. Thus, lot sizing decisions taken at time t cannot affect the backorders or lost sales arising over the frozen part of the schedule, periods $t+1$ to $t+F$, but start only for orders that will arrive in periods $t+F+1$ onwards. The problem is therefore to find the orders from this period $t+F+1$ onwards that minimise the costs of meeting the forecast demand over the next $F+P$ periods where P is the free planning horizon. Let $F_{t,t+k}$ be the forecast of demand based on the forecasting method in use (to which is added any desired buffer) for period $t+k$ made at the end of period t , after the demand for period t is known, for $k=1$ to $F+P$. There are outstanding orders from earlier periods, OO_{t+k} , scheduled to arrive for period $t+k$ for $k=1$ to F . Let $I_{t,t}$ be the known stock at the end of period t . $I_{t,t+k}$ is defined as the stock projected for the end of period $t+k$ on the basis of the forecast period demands, outstanding orders and the ordering decisions made in period t .

Different service levels are generated by specifying the target stock levels to be achieved at the end of every period including buffers. Unlike the standard lot sizing procedure for deterministic demands, the aim here is to ensure that the projected start of period stock, plus the order scheduled to arrive at the start of that period, less the forecast demand is at least some specified stock (SS) level. If SS is positive then it is a safety stock to guard against demand variations. If SS is negative, then it is the maximum level of backorders that can be planned to be carried from one period to the next or, in the lost sales case, the maximum shortfall in meeting the predicted demand for a period. So in the lot size planning in period t , new orders should be scheduled to arrive in period $t+k$, $O_{t,t+k}^*$, to ensure that:

$$I_{t,t+k} = I_{t,t+k-1} + OO_{t+k} + O_{t,t+k}^* - F_{t,t+k} \geq SS$$

for $k = F+1$ to $F+P$ (6)

treating the forecasts as realised demands. The lot sizing problem at the end of period t can thus be expressed as to determine the orders scheduled for arrival in period k , $O_{t,t+k}^*$, that minimise the total costs of ordering and stockholding over periods $t+F+1$ to $t+F+P$ whilst satisfying the above constraint for all periods. A fuller discussion is given, for example, by Jeunet (2006) or Voss and Woodruff (2006). For the backorders case see Fildes and Kingsman (1997).

3.5. Lot sizing rules (LSRs)

Six alternative lot-sizing methods were evaluated in the simulation, as given below, and include those commonly adopted in organisations. Fuller descriptions are given in texts such as Orlicky (1975) and Silver *et al* (1998). In all cases, the currently uncovered net requirements over the free part of the planning horizon, taking account of any backorders that have accumulated, are firstly determined to give the demands to be satisfied. The average period demand is the average of the uncovered net requirements divided by the total number of periods in the free part of the planning horizon. However, if the net requirement in the first free period (or first few free periods) is zero then this is ignored in the averaging. In this particular case, no order will be placed and the simulation moves on to the following period, updating the system with the actual period demand.

Economic Order Quantity (EOQ): The lot size is set equal to the standard Wilson Economic Order Quantity calculated on the basis of the average period demand, as above. If the EOQ is insufficient to meet the first period's net requirement, the lot size is made the smallest multiple of the EOQ that achieves this.

Period Order Quantity (POQ): This converts the EOQ lot size to the nearest integer number of periods demands, (again based on the average period demand over the free part of the planning horizon, as above).

Least Unit Cost (LUC): The lot size is increased to cover succeeding periods until the item unit cost is at a minimum. The item unit cost is the ordering cost plus the stockholding costs for the lot divided by the lot size. If there is a tie, then the smaller number of periods is taken.

Least Total Cost (LTC) or Part Period Balancing: This is based on the concept that the least total cost occurs where the ordering and stockholding costs are most equal. The lot size is increased to cover succeeding periods until the stockholding cost is closest to the ordering cost. Again, if there is a tie, the smaller number of periods is taken.

Silver Meal Heuristic (SM): The objective is to minimise the costs per period, so the total costs are divided by the number of periods. If a period within the total covered by the lot has zero net requirement then it counts as a period for the division.

Wagner-Whitin algorithm (WW): This procedure gives the optimal solution for minimising the costs of meeting a

given set of deterministic requirements over a given planning horizon.

In addition, the simulation included the Lot-for-Lot method and the incremental heuristic, mainly as a check of the accuracy of the simulation programme (Silver *et al.*, 1998). However, the results for these methods are excluded from the analysis due to their consistently poor performance.

4. The simulation experiment

Within the production system described in the previous section, a number of arbitrary decisions have been made that could affect our results. Demand has been generated over 20 000 periods but the variability of demand itself suggests it is wise to repeat the simulation experiment in order to establish the effect on demand, service level and the various costs. Five replications were carried out and detailed examination analysis of these runs showed that there was very little difference between the outcomes of each of the replications. The safety stock was set as a buffering factor times the forecast error standard deviation (FORDEV). An extended range of buffering factors were chosen to ensure sufficient data over service levels above 90% to deliver a good approximation to the cost-service relationship. The service level recorded was the average value over the five replications for that buffering factor, and similarly for the unit costs. The buffering values were chosen so as to cover service levels in the range 90% to 99.5%.

The random numbers used were generated using the Turbo Pascal function 'random' based on fixed seeds, so that each experimental run used common random numbers to reduce variance (Law and Kelton, 2000). The normal distribution generator was that proposed by Box and Müller (1958).

In addition the effects of the following parameters on the results have been examined:

- (1) *Frozen interval*—the experiment was run with three different frozen periods: 2 (the minimum possible given the two period nature of the production process), 5 and 7. The results reported in this paper are only for a frozen period of 5 to conserve space. The results for the other frozen intervals are similar. Unsurprisingly, the shorter the frozen interval, the lower the unit cost curve. The ranking of the LSRs remained essentially unchanged (average rank correlation over five replications = 0.97). The change in the average unit cost is small, averaging around 1%.

Because the frozen interval determines the first period in which the schedule can be changed it also affects the actual error standard deviation under which the plans are set. The error standard deviation for the first period in which the schedule can be changed, denoted by $\sigma(F+1)$ where F is the length of the frozen interval in periods,

for both demand Data Generating Processes has been set at the values 50, 100 etc using Equation (3). (Only in the case of the autoregressive DGP does this lead to a rebasing of the error standard deviation; for the Random Noise Process the lead-time error standard deviation remains constant whatever the lead time.)

- (2) *Planning horizon*—the simulations were run with three different free planning horizons, 12, 24 and 36 periods, for some options. Note that the full planning horizon is made up of the relevant frozen interval plus this free horizon. The results were virtually identical with an average correlation of 0.99 approximately. This was expected because even the smaller free planning horizon is more than twice the maximum TBO of between 4 and 5 periods. All the experiments reported are for a free planning horizon of 24 periods except where noted for the Wagner–Whitin algorithm.
- (3) *Simulation run length and warm-up period*—the experiment starts with zero stocks and zero demand over the frozen interval. The actual simulation period was for 20 000 periods, having first 'run-in' the simulation for a warm-up period of 72 periods. This is sufficient for the frozen period plus the planning horizon considered here. The warm-up period outcomes are excluded from the analysis of the results.
- (4) *Verification*—the programme was built up from a number of test sub-programs covering the lot sizing rules, the demand and forecast generation processes and the rolling schedule calculations. Each was verified against manual calculations.

5. The effects of demand uncertainty with optimal forecasting

The results of the various simulations are presented in terms of two key measures of performance as discussed earlier; the mean unit cost and the mean service level for each combination of experimental factors, averaged over the $J = 5$ simulation replications.

Experiments were performed on two autoregressive demand processes, the Random Noise DGP, $AR(\rho = 0)$, and $AR(\rho = 0.9)$.

Figure 1(a) presents some examples of the trade off curves obtained for the various lot sizing rules with a Random Noise DGP which show the mean unit cost graphed against the mean stock-out (shortage) level (between 0% and 8%) for each of the buffer values used in the experiments when there is perfect foreknowledge of the future demands. This is the deterministic demand case for which the LSRs were originally designed. The results shown are for a TBO value of 4 and for two values for the error standard deviation, 100 and 400. (It should be remembered that a $TBO = 4$ means that the average time between orders is 4 periods, which is given by an ordering cost of 7000 relative to a period holding cost of 1).

Figure 1 Unit cost *versus* shortage level various lot sizing rules: time between orders 4 periods. (a) Perfect information. (b) Uncertain demand.

A tentative finding from Figure 1(a) is that there is little difference between the unit costs of the different LSRs for the lower error standard deviation, apart from the EOQ, which costs significantly more. The relative ordering of the unit costs of the LSRs remains the same over the whole range of service levels shown. The differences between the rules become larger as the error standard deviation increases. The ordering of the rules in terms of unit costs, lowest cost first, is WW, SM, LTC, LUC, POQ and EOQ. However, these results give a base level for comparison of the effects of demand uncertainty on the performance of the rules. The WW rule, of course, gives the optimal solution over the planning horizon at each period in turn, allowing for the previous decisions although this is not assured for a finite planning horizon (sensitivity tests showed it optimal for a horizon of 90). The EOQ performs worst because its order size is not constrained to be the demand over a number of integer periods, leading to carry-over stocks in every period. This is not new, being typical of what has been found in previous studies (see Silver *et al.*, 1998, for a summary and references on comparative lot sizing performance).

Figure 1(b) shows the trade off curves obtained for the various lot sizing rules for the same case as in Figure 1(a) but now based on forecasts of the future demands each period. The ordering decisions are based on the forecast of demands

over the planning horizon at each period in turn, not the actual demands as for Figure 1(a). Note that in this case, the unit costs become asymptotically large as the service level tends to 100%. An immediate finding from Figure 1(b) is that there is again a consistent difference between the different LSRs over the whole range of service levels shown, but that the differences are much smaller than those of Figure 1(a) (the deterministic demand situation with perfect foresight). The relative ordering of the unit costs of the LSRs is the same for both levels of demand uncertainty but the differences become smaller as the noise standard deviation increases. The other preliminary finding is that ordering of the rules in terms of unit costs is the reverse of what was found for the perfect foresight case. The EOQ is always the best, whilst the WW 'optimal' algorithm is almost always the worse. However, there is not a lot of difference between the rules other than for the EOQ.

Graphs such as Figure 1 are an illuminating way of presenting complex results, but, unfortunately, the service levels resulting for each lot sizing rule and level of demand uncertainty for the same buffering factor (and hence safety stock) differ. Although one can draw general conclusions from such figures, they do not allow the direct derivation of a quantitative summary of the relative performance for all costs structures and the different levels of the error variance

Table 2 Ranking the rules by unit cost (measured by ranking each method for each TBO and level of noise uncertainty and summing the ranks)

<i>Random Noise data generating process</i>		<i>Autoregressive demand DGP</i>	
<i>Deterministic demand–perfect information</i>	<i>Forecasting demand</i>	<i>Deterministic demand–perfect information</i>	<i>Forecasting demand</i>
WW	EOQ	WW	LTC
SM	LUC	SM	EOQ
POQ, LTC	LTC, POQ, SM	LUC	POQ ,
LUC		LTC, POQ	
EOQ	WW	EOQ	SM, WW, LUC

for a particular DGP. The proposed solution to this difficulty, new to this literature, is to model the trade off curves quantitatively to give the relationship between the unit cost and the service level (see Johnston *et al.*, 1988) who propose a similar concept. For each lot sizing rule, error standard deviation, data generation process and cost structure (*TBO*), the logarithm of the unit cost for the *i*th experimental condition was taken to be a function of the out-of-stock level (equal to 100-service level):

$$\begin{aligned} \log_e(\text{Unit Cost}_i) &= \mu + \beta_1(100 - \text{Service Level}_i) \\ &+ \beta_2(100 - \text{Service Level}_i)^2 \\ &+ \beta_3(100 - \text{Service Level}_i) + \beta_4(100 - \text{Service Level}_i)^3 \\ &+ \log_e(100 - \text{Service Level}_i) + \varepsilon_i. \end{aligned} \tag{7}$$

Various simpler alternative specifications were considered such as a log-log function of mean service level but this more flexible form was preferred as it gave a consistently better description of the complex cost responses to the different experimental conditions. The curves were estimated for service levels greater than 90%. The maximum percentage error in predicted unit cost is less than 0.1% with $R^2 = 1$ (approximately) in all cases. Different plausible functions led to the same rankings of the various LSRs. The estimated parameters depend on such aspects of the simulation experiments as the frozen period, planning horizon etc.

Using these fitted curves, it is then possible to estimate the unit costs for each lot sizing rule, *TBO* and error standard deviation for any particular service level and so perform a quantitative analysis of relative performances. Two target service levels were selected, 99% and 96%, representing high and a more modest but acceptable level of service achievement.

An analysis of ranking the lot-sizing rules performance by unit cost for each of the *TBO* and error standard deviation combinations confirmed the earlier result on their relative ordering, see Table 2. The ordering is as expected when there is perfect information on future demand for both DGPs. The optimal WW rule is first followed quite closely

by SM. The EOQ is always worst by far. The order of the other three rules varies between the two DGPs. The ordering is reversed when we have to forecast the demand using the ‘optimal’ forecasting method for the particular process. The WW rule, the optimal for the deterministic situation is now very much the worst rule. The EOQ, the worst by far in the deterministic case, is best for the Random Noise DGP and second best for the Autoregressive DGP. The LTC rule, which was poor in the deterministic case, is now the best for the Autoregressive DGP. Thus having to forecast demand and plan on the basis of a rolling horizon in a stochastic demand situation gives very different results to those found under the deterministic demand situation for which the LSRs were devised.

To assess the implications of these results we need firstly to check if there is a significant difference between the performance of the rules. In order to avoid any assumption of normality, a non-parametric test, the Friedman Test for the two-way analysis of variance by ranks, was performed to test if there are significant differences in performance between the six lot sizing rules and the effect of varying *TBO*s and error standard deviation. The use of common random numbers, while introducing dependence between unit costs for different treatments, does not undermine either Friedman’s test or the ANOVA tests considered later as the comparisons are blocked (as in a paired *t*-test). The *p*-values of the Friedman statistic indicated that overall there are significant differences in performance (at 5%) between the different lot sizing rules for the Random Noise Process but not for the Autoregressive Process. There is somewhat more variation in performance across *TBO*s compared to the error standard deviations, though the differences are not pronounced.

Another way to measure the relative performances is to calculate the ‘regret’ of using a particular lot sizing rule, that is, the increase in unit cost compared to the best method for that *TBO*, error standard deviation and service level. The overall average regret values for each of the lot sizing rules for each service level for the Random Noise and Autoregressive Demand DGPs are displayed in Table 3. This confirms the earlier deductions above.

It can be seen for the Random Noise DGP that the EOQ rule always gives the best performance by far for both service

Table 3 The overall average percentage regret by error standard deviation by Lot Sizing Rule

3a) Random Noise DGP

<i>Lot sizing rules</i>	<i>Service level</i>							
	96%				99%			
	<i>Error standard deviation</i>				<i>Error standard deviation</i>			
	50	100	200	400	50	100	200	400
EOQ	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
POQ	6.8	4.9	2.4	0.9	4.5	3.9	2.0	0.5
LUC	6.6	4.0	1.3	0.3	5.2	3.9	1.5	0.3
LTC	6.8	4.9	2.4	0.9	4.5	3.9	2.0	0.5
SM	6.8	4.9	2.4	0.9	4.5	3.9	2.0	0.5
WW	6.8	4.9	2.4	0.9	4.6	4.0	2.0	0.6

	TBO				TBO			
	2	3	4	5	2	3	4	5
	EOQ	0.0	0.1	0.0	0.0	0.0	0.0	0.0
POQ	8.4	2.9	2.1	1.5	5.9	2.1	1.7	1.3
LUC	5.9	3.3	1.8	1.1	4.8	2.9	1.8	1.3
LTC	8.4	2.9	2.1	1.5	5.9	2.1	1.7	1.3
SM	8.4	2.9	2.1	1.5	5.9	2.1	1.7	1.3
WW	8.4	2.9	2.1	1.6	5.9	2.1	1.7	1.5

3b) Autoregressive (0.9) DGP

<i>Lot sizing rules</i>	<i>Service level</i>							
	96%				99%			
	<i>Error standard deviation</i>				<i>Error standard deviation</i>			
	50	100	200	400	50	100	200	400
EOQ	0.0	0.0	0.7	2.2	0.0	0.0	0.8	1.5
POQ	4.0	1.8	0.5	0.7	3.1	1.5	0.2	0.5
LUC	3.3	0.9	0.7	2.2	3.2	1.1	0.7	2.1
LTC	4.0	1.7	0.2	0.0	3.1	1.5	0.0	0.1
SM	4.0	1.8	0.7	1.2	3.1	1.5	0.5	0.7
WW	4.8	2.5	1.3	1.4	3.8	2.1	0.9	0.9

	TBO				TBO			
	2	3	4	5	2	3	4	5
	EOQ	0.4	1.2	0.7	0.6	0.8	0.6	0.2
POQ	4.5	1.2	0.7	0.5	0.7	0.6	0.4	0.5
LUC	3.5	2.0	1.0	0.7	1.0	0.9	0.5	0.8
LTC	4.3	0.9	0.5	0.2	0.3	0.3	0.3	0.3
SM	4.5	1.7	0.9	0.6	0.8	0.7	0.5	0.5
WW	4.5	1.6	0.7	3.3	3.1	2.9	2.8	2.6

levels, with a regret of very close to 0%. The LUC rule is slightly better than POQ, LTC and SM rules, which are again slightly better than the WW rule. The size of the regret falls as the error standard deviation increases and also as the *TBO* takes higher values. The potential saving in choosing the EOQ rule compared to any other is around 8% for a *TBO* of 2 but falls to only 1.4% for the highest *TBO* of 5, for the 96% service level. For the lowest error standard deviation the regret is 6.4% falling to only 1.4% for the highest value of 400.

The impact becomes smaller as the service level increases from 96% to 99%. To summarise, the choice of lot-sizing rule becomes less critical as the demand uncertainty increases and as the average time between orders increases.

Comparing Table 3b with 3a shows that the regrets from using the best rule in each case are smaller for the Autoregressive DGP than from the Random Noise DGP. The EOQ is best for the lower error standard deviation whilst the LTC rule is best for the two higher error standard deviation. The

Table 4 The percentage increase of best LSR rule above the WW solution based on perfect demand information

4a) Random Noise DGP

<i>TBO</i>	<i>Service Level</i>							
	96%				99%			
	<i>Error standard deviation</i>				<i>Error standard deviation</i>			
	50	100	200	400	50	100	200	400
2	15.8	20.7	40.3	102.9	26.5	35.6	64.2	146.3
3	7.9	10.9	22.8	60.3	15.1	20.6	38.5	88.6
4	4.1	6.0	13.2	37.2	9.1	12.6	23.8	56.1
5	2.3	3.6	9.0	26.3	6.3	8.6	16.9	40.6
Average	7.5	10.3	21.3	56.7	14.2	19.3	35.8	82.9

4b) Autoregressive (0.9) DGP

<i>TBO</i>	<i>Service level</i>							
	96%				99%			
	<i>Error standard deviation</i>				<i>Error standard deviation</i>			
	50	100	200	400	50	100	200	400
2	18.3	31.1	71.4	180.0	31.7	51.6	106.6	243.2
3	9.1	15.9	38.0	108.1	17.6	29.1	60.6	150.6
4	4.6	8.5	22.7	67.7	10.6	17.4	37.8	97.1
5	2.7	5.5	15.4	48.1	7.2	12.4	27.1	69.9
Average	8.7	15.3	36.9	101.0	16.8	27.6	58.0	140.2

EOQ rule is best for a *TBO* of 2 but the LTC takes over for higher *TBO*s of 3, 4 or 5. The autoregressive results show that generally the percentage regret from using the wrong rule is less than 2%. Thus unlike the Random Noise DGP, the choice of LSR to use for the Autoregressive DGP is unimportant except for very low (and unrealistic) error standard deviation and *TBO*s.

A possible explanation of the relative performance of the various lot sizing rules is that for a given set of forecast demands, the standard lot sizing heuristics cumulate demands over an integer number of periods. This avoids the deliberate creation of carry-over stock as occurs with the EOQ rule. The differences between the forecasts and the actual demands will create unplanned carry-over stock and also periods with small net requirements. Thus, the existence of demand uncertainty negates the advantage of the rounding to match period demands of the lot sizing rules in the deterministic demand situation. The EOQ rule carries an implicit safety stock also. Consequently, when the final product demand has to be forecast, using the EOQ for lot sizing works effectively. Once forecast errors are acknowledged, there will always be carry-over stock from period to period, so any method that concentrates on minimising costs is perhaps not as effective as examining costs per unit acquired.

To summarise the results, the choice of lot-sizing rule, for both processes, becomes less critical for increases in demand uncertainty, the average time between orders and the service

level, and becomes close to inconsequential for higher *TBO*s and service levels.

It is also valuable to assess the increase in cost arising when forecasting uncertain demand compared with the costs resulting from known demand. This is illustrated in Table 4 where costs are compared to those calculated with perfect foresight of the (stochastically) generated demand pattern for both DGPs. This shows the percentage increase in unit cost for the best lot sizing rule above the unit cost for the corresponding WW cost, based on the deterministic case of perfect foresight (disaggregated for the different *TBO*s and error standard deviations). The bottom row of the Table shows the average increases over *TBO* values from 2 to 5. The increases are much higher for the 99% service level than for the 96% service level. The values derived indicate a complex pattern of behaviour, with the standardised unit cost increase rising as the time between orders lengthens, the error standard deviation increases and the service level increases. There is again an economically significant difference between the two DGPs. The cost increases due to the uncertainty in the demand are much greater for the Autoregressive DGP than for the Random Noise DGP. The average of the 96% and 99% service levels can be used as an overall indicator of the effect of changes in the error standard deviations. The rate of increase in unit cost as the error standard deviation increases is approximately exponential, at 16% and 29% for the Random Noise and the Autoregressive DGP, respectively.

These results contradict De Bodt and Van Wassenhove's (1983a) often repeated conclusion that with any level of demand uncertainty greater than zero, costs were increased by about 10%. In fact, their result is an artefact arising from the attempt to eliminate service level as a factor in their experiments. This more complex and conditional result points to the circumstances where uncertainty matters most: particularly when the time between orders is short, service levels are high and the demand is positively autoregressive. Essentially, the ordering processes allow errors to mount up with a positive autoregressive process in contrast to a negative autocorrelation process.

6. The value of forecasting and the effects of mis-specification

Now the above results are indicative of the value of forecasting. Transparently, changes in the demand uncertainty leads to significant changes in the unit cost when service level is kept fixed. However, this does not directly answer the question posed in the introduction: what is the value of forecasting? Whilst to most practitioners, the forecast error is the difference between the actual and the forecast value, as we pointed out in Sub-section 2.3, this error combines the randomness in the process generating the demands and the errors arising from poor forecasting. The overall error is the difference of the two errors (see Equation (5)) and assuming, as we do, that the process and forecast errors are uncorrelated, its variance is the sum of the two variances. It is only the forecast error component that represents the potential value of improving forecasting accuracy.

Figure 2 shows the effects on the unit cost/shortage level trade-off curves of increasing the forecasting error by 20% and 50% compared to when the forecast is optimal (ie $v_f \equiv 0$). This is further contrasted with the curve based on perfect foresight information. The illustrative graph of the results shown is derived using the EOQ lot sizing rule and for a TBO of 3 periods for different standard deviations in the data generation process. The picture remains the same for the

various TBOs. It shows the substantial increase in unit cost (for a given service level) as we move from perfect information through the different levels of forecast error, and in particular, the proportionate benefits across all service levels of improved accuracy.

In order to quantify the value of forecasting, the logarithm of unit costs has been modelled as a function of possible explanatory factors as given in Table 5. As before, the model was of the \log_e (average unit cost) form. The data set modelled was for mean service levels lying between 93% and 99.9% to avoid distortions at the extremes of the data. This range includes those values of practical interest.

In the model specification the service level has been included in the same form as was used to estimate the 99% and 96% service levels in Section 5, Equation (7). The results show that the factors LSR and Cost (as measured by TBO) are significant. Unsurprisingly, the largest reduction in variance is due to the Cost (TBO) parameter. The noise standard deviation of the data generation process, *STD*, and the forecast error standard deviation (*FORDEV*), were included linearly. The inclusion of the variances STD^2 and $FORDEV^2$, as well as the standard deviation lead to improved predictions for the larger values of the overall deviation. The RMSE is low suggesting a satisfactory model (with inter-quartile ranges of approximately $\pm 5\%$ for the error) with R^2 of 99%. Details of the associated analysis of variance summary table are omitted as the inevitable consequence of using such a large data set is that almost all factors are found to be 'significant'. Almost all earlier research has usually confused 'significance' with economic importance whilst what managers need to estimate is the latter.

Graphical analysis plus the inclusion of additional cross-product terms such as the often insignificant interactions between the overall error standard deviation and LSR and TBO suggested that the model was broadly adequate and the error distribution was approximately normal but it failed to capture all the non-linearities induced by the interaction of *STD*, the standard deviation of the DGP, and the cost structure, TBO. However, including such interaction terms

Figure 2 Unit cost versus shortage level as it is affected by forecast error levels based on EOQ: time between orders of 3 periods.

Table 5 Possible explanatory factors in a model of $\text{Log}_e(\text{Cost})$.

<i>Qualitative factors</i>	<i>Quantitative factors</i>
Lot sizing rules (LSR) Cost structure (TBO)	Noise standard deviation of the DGP: (<i>STD</i>) and its variance, STD^2 Standard deviation of the forecast uncertainty—measured as <i>FORDEV</i> and its variance, $FORDEV^2$ An interaction term, $PROD = STD * FORDEV$
Forecast error level (FORDUM) Interaction of (LSR and TBO) and (LSR and <i>STD</i>)	Overall error standard deviation, $OVERALLDEV = \sqrt{STD^2 + FORDEV^2}$ The out-of-stock level, <i>SERV</i> —measured as: (100-Service Level), and transformed as Equation (4).

Table 6 Testing the difference between Lot Sizing Rules—the contrasts

<i>Contrast</i>	<i>Estimated effect (% difference in unit cost)</i>		
	<i>All data</i>	<i>No forecast error</i>	<i>With forecast error</i>
EOQ <i>versus</i> POQ	-0.74**	-2.80**	0.09
EOQ <i>versus</i> LUC	-0.80**	-2.65**	-0.06
POQ <i>versus</i> LUC	-0.06	0.15	-0.15
EOQ <i>versus</i> LTC	-1.28**	-2.80**	-0.61**
EOQ <i>versus</i> SM	-0.80**	-2.80**	0.00
EOQ <i>versus</i> WW	-0.84**	-2.85**	-0.03
EOQ/LUC <i>versus</i> rest	-0.52**	-1.50**	-0.11
EOQ/LUC/ POQ <i>versus</i> rest	-0.69**	-1.51**	-0.33**
SM <i>versus</i> WW	-0.04	-0.06	-0.03

**Represents a contrast significant at less than 1%.

has no effect on the qualitative conclusions we draw and little effect on the quantitative estimates we make since the coefficients are small. The final specification adopted included a dummy variable to distinguish between the situation with zero *FORDEV* ($FORDEV = 0$) equivalent to optimal forecasting and non-zero forecast error. Its interaction with LSR was also included.

The key interaction to include is that between the cost parameter and the overall uncertainty, preferred to both the process *STD* and the *FORDEV*. This shows that the impact of the overall deviation lessening as the time between orders (*TBO*) increases. Essentially the shorter the time between orders the more sensitive the system is to demand uncertainty and forecast error. The impact of the *FORDEV* is larger, compared to the process error, though there is a qualitative reduction in cost in moving from the zero *FORDEV* to the non-zero case, ie small amounts of forecast error lead to overall improvements in performance.

The above model was used to evaluate the performance of the different lot sizing rules on unit cost. Overall EOQ, POQ and LUC performed best (although the estimated cost difference is small, typically less than 1%) as shown in the column ‘Estimated Effect’ in Table 6. However, this overall result disguises a more complex picture that becomes apparent if the analysis is carried out only for those points with zero *FORDEV* on the one hand and positive deviation on the other. While overall EOQ is not the uniform best performer,

for zero *FORDEV* it is significantly better than its competitors (confirming Tables 2 and 3) whilst the best performers with forecast error were EOQ, POQ and LUC which differed significantly from the rest. Whatever the error standard deviation, there was no significant difference found between Silver-Meal and Wagner-Whitin.

Estimating the value of forecasting requires some assumptions about the size of the overall uncertainty that arises in practice and also how it is divided between the noise in the process generating demand and the forecast errors incurred from using a non-optimal forecasting method. Only the second source of error is lessened by improved forecasting. It was stated in Section 2 that, in practice for short term forecasting, weekly or monthly up to three steps ahead, the standard error of the forecast for faster moving items would not usually exceed 40% of the mean demand. Improvements in forecasting of around 20–50% in the overall forecast error are possible by implementing appropriate forecasting techniques to match the particular characteristics of the demand data used in the MRP system (Fildes *et al*, 1998, 2009). Thus, it would be indicative of possible savings in unit cost to consider the improvements expected for values of 200 and 400 in the overall uncertainty, derived from different mixes in process noise and forecast error. Using the above response model specification in Table 5, the estimated percentage change in cost is independent of cost, LSR and service level. Using a first order Taylor’s approximation for the response function

Table 7 The effects of forecast model mis-specification.

		Data Generating Process		
		Forecast noise	AR(0) (%)	AR(0.9) (%)
Forecasting Generating Process	AR(0)	0%	15.4	46.1
		20%	26.4 Quadrant I	50.0 Quadrant II
		50%	39.3	54.8
	AR(0.9)	0%	25.6	25.1
		20%	67.6 Quadrant IV	46.7 Quadrant III
		50%	107	67.4

Note: % loss compared with perfect information, averaged over the ‘better’ lot sizing rules, EOQ, POQ and LTC, averaged over *TBOs* 3 & 4, for process uncertainty of 100 and 200 and both service levels.

as a function of *h*, the percentage improvement in unit cost for an *h*% improvement in forecasting accuracy is given by:

$$\left(\beta_{TBO,OVLDV} \frac{FORDEV}{OVERALLDEV} + \beta_{fordev} + 2\beta_{fordevsq} FORDEV + \beta_{prod} STD \right) \times FORDEV \times h \tag{8}$$

where *FORDEV* is the forecast error, and *OVERALLDEV* is the overall uncertainty. The β coefficients have the obvious meaning with $\beta_{TBO,OVERALLDEV}$ representing the interaction term. In quantitative terms β_{FORDEV} dominates the cost improvement from improved accuracy but the formula shows the size of the forecast error, measured by *FORDEV* × *h* is also critical. The approximation underlines the importance of the *FORDEV* relative to the process noise. In effect major improvements in cost can be achieved where the overall uncertainty is high and a major component of this is forecast error. However the table highlights the difficult of making major gains. The conclusions from the AR(0.9) analysis were similar with a higher sensitivity. An alternative approach to the problem of estimating the value of improved forecasting accuracy is by using the raw simulation results and comparing the unit cost of dropping from 50% forecast error to 20% to 0% for a fixed service level. This approach gives higher estimates of the forecast sensitivity with median forecast sensitivity at the 96% service level of 11% and 19% for demand uncertainty of 200 and 400, respectively. A higher service level of 99% adds another 2 percentage points. For the AR(0.9) process the effects of forecasting improvement are uniformly larger with approximate unit cost improvements of around 6.5 percentage points above the AR(0) benchmark values given above.

6.1. Mis-specification

In practice, simple forecasting methods are usually applied without any prior analysis of the appropriate underlying model of the DGP. The effect is to introduce both forecast error and mis-specification error. As noted in the introduction

such mis-specification is a common feature of various supply chain models discussed in the literature including recently, for example, Chen *et al* (2000a) and Zhao *et al* (2001). A full discussion of its effects would ensure the paper was overly long. However, we have considered two cases, where the DGP is the Random Noise model and the forecasting model is the AR(0.9) and vice versa. Table 7 gives the results averaged over the two service levels and lot-sizing rules EOQ, POQ and LTC which ensured the ‘better’ LSRs have been used.

Table 7 illustrates the effects of using a mis-specified model (typically the case in previous research) compared with using an optimal model for the particular case. The table is split into four quadrants, the first under the heading AR(0) repeats the unit cost where the DGP is an AR(0) process and the optimal forecasting model specification has been used to model the DGP (but with three levels of forecast error, 0, 20 and 50%, of the demand uncertainty). The fourth quadrant shows the results where the DGP is again the AR(0) but the forecasting model that is used is AR(0.9) (again with the three levels of noise). Quadrants II and III are similar but the DGP is AR(0.9), so for example, quadrant II shows the results of using a Random Noise forecasting model when the true (but unknown) DGP is autoregressive. Thus there is a loss of 46% if a Random Noise model is used when the true process is autoregressive compared to 15.4% if the correct model had been correctly used.

The results strongly confirm the lack of symmetry in the two specification errors in that the regret from choosing the Random Noise Process decreases as the forecast noise increases, but the reverse is true if the autoregressive process is chosen. Ideally, whatever the level of forecast error and DGP the forecaster would prefer to know that the forecasting model he uses is better than the alternatives. However, here no dominant strategy is available for selecting a forecasting model although choosing the simpler AR(0) model seems the more effective for higher levels of forecast error. An analysis of the more detailed results show the specification error effects on costs increases with *TBO*, process uncertainty and service level.

These illustrative results suggest that earlier research studies (which make assumptions equivalent to mis-specifying

the forecasting model adopted) are likely to have mis-stated the effects of demand uncertainty, potentially substantially, and their conclusions will be dependent on the nature of the mis-specification assumed in the research design.

7. Discussion and conclusions

In this research we have demonstrated a methodology for integrating aspects of manufacturing research with forecasting. The methodological issues we have explored aimed at highlighting deficiencies in previous research designs: the conflation of demand uncertainty with forecasting error uncertainty and model mis-specification, the use of 100% service levels to draw conclusions and the pervasive use of analysis of variance when the issue of significance can easily be dealt with through increased sample size, thus neglecting the much more fundamental question of the practical importance of the results. The size of key endogenous variables (eg the effects of forecast error on cost) is important, when choosing a forecasting system or developing an information sharing arrangement as in other policy areas (Ziliak and McCloskey, 2004). The substantive aim of the research has therefore been to show how to estimate the value of improved forecasting and to flesh out the claims made by software suppliers and forecasters alike that accurate forecasting is critical to many manufacturing operations.

Various stylised facts spring from the substantive analysis we have presented.

- Unit cost (expressed as a percentage of the unit cost based on perfect foresight) increases exponentially with demand uncertainty in contrast to De Bodt and Van Wassenhove (1983a) assertion. The discrepancy with this study apparently arises from their focus on service levels of 100% only.
- In the more general framework established here, the best lot sizing rules when demand uncertainty exists are very different to those based on deterministic demand which assumes perfect foresight. This strengthens, for example, Wemmerlöv and Whybark's (1984) conclusions. For lot sizing research to make a contribution to improved manufacturing, uncertain demand must be explicitly considered in a realistic context that includes a variety of experimental factors, in particular service level. The benefits of choosing the 'best' rule are typically low. Applying the results from deterministic lot sizing research is clearly inadequate (Wemmerlöv, 1989). The focus in earlier research on 100% service gives misleading results as to the effects of uncertainty and forecast error on unit cost and service level. (Previous results on comparative lot sizing hold true however.)
- The benefits of improved forecasting increase with overall uncertainty but this depends on the relative sizes of the stochastic variation in the demand generation process and the forecasting errors. These benefits are substantially greater than the effects arising from the choice of lot sizing

rule. Potential savings are over 10% for higher levels of demand and forecast uncertainty and realistic levels of forecast improvement. Sensitivity to forecast error is higher for the autoregressive process.

- Mis-specification in the choice of a forecasting model leads to increased forecast error and increased costs. There is no best method of forecasting; it will in general depend on the 'true' but unknown demand generation process. There is no dominant strategy of forecasting model selection as measured by regret, despite the common practice in both research and organisational forecasting practice of using a generic model such as exponential smoothing in all applications. The benefits of forecast model selection will depend on the noise in the data.

Taking the results as a whole, they show that while it is sometimes important to consider the choice of lot sizing rule, attempts to reduce the forecast errors are likely to prove more valuable, as the lot sizing choice is relatively straightforward depending on *TBO*. As discussed in Section 5, the EOQ is the best choice for the Random Noise Process but LTC performs somewhat better for the autoregressive DGP. This reduction in forecast error can be achieved both by moving closer to the optimal forecasting method for the problem and also by trying to reduce the variation in the data generation process itself. There are, however, developments in lot sizing that recognise the stochastic nature of demand (eg Babai and Dallery, 2009) which could have different performance characteristics.

The value of improvements in forecasting depends quite critically on the cost structure of the manufacturing process. The emphasis in recent developments in production and operations management, whether called achieving world class manufacturing, lean production methods or Just In Time (JIT) production, has been on the benefits of reducing lot sizes, so the values for lower *TBOs* may be better indicators of the impact of demand uncertainty. If manufacturing is improved to make lower lot sizes more economic for production, then this results in the costs of satisfying customer demands being even more sensitive to uncertainties in demand, whether due to inherent process variation or to inadequate forecasting. Thus, it is important to manage the demand generation process as well as to take actions to make lower lot sizes more economic in order to improve the overall performance.

A number of substantive research questions remain, in particular whether the results generalise to other (more complex and realistic) supply chain processes, in particular through the inclusion of capacity constraints, correlated demand and JIT procedures. Case research is needed here. In addition, little of the research done on forecasting has addressed the types of data generation processes observed in manufacturing systems which may well be severely non-normal, perhaps with intermittent demand, outliers and changes in trend (Gardner, 1990; Fildes *et al.*, 2009). The importance of this issue, totally neglected in earlier lot sizing research, has been demonstrated in the results for the more

realistic autoregressive process, which show a higher sensitivity to forecast error. We would expect that the types of specification error commonly arising when automatic forecasting systems are used, would lead to increased advantages to more accurate, properly specified forecasting models. The key question remaining is what improvements in forecasting accuracy are realistically achievable. As this paper has shown, the benefits to success in achieving such improvements are worth the effort.

Postscript: The last twenty years has seen increasing research on uncertainty in supply chain planning, particularly on aspects of the problem such as the bullwhip effect. The primary emphasis in most of the research has been on mathematical tractability and ‘significance’. It has also failed to distinguish between forecast error and uncertainty. The result has been few findings of any managerial importance. For example, we may not be too surprised to know that collaboration through information sharing between retailer and supplier is sometimes valuable. Worse still, some of the general conclusions such as the inevitability of the bullwhip effect are wrong, the product of an inadequate or overly specific research design. This failure of the research community to combine practically valuable modelling of the supply chain with high profile academic research is important to all researchers who face increased pressure to publish in the ‘top journals’. These journals are dominated by research paradigms that emphasize mathematics at the expense of relevance and grounding in observed practice. In short, more practice, more empirical modelling that includes forecast error -and less reliance on spurious mathematical simplicity are required. It is a call that OR researchers have heard since OR’s founding as an academic discipline in the 1950s. And it is still worth repeating. I’m confident that Brian Kingsman would have agreed with me.

References

- Aviv Y (2001). The effect of collaborative forecasting on supply chain performance. *Mngt Sci* **47**: 1326–1343.
- Aviv Y (2002). Gaining benefits from joint forecasting and replenishment processes: the case of autocorrelated demand. *Manuf Serv Ops Mngt* **4**: 55–74.
- Babai MZ and Dallery Y (2009). Dynamic versus static control policies in single stage production-inventory systems. *Int J Prod Res* **47**: 415–434.
- Box GEP and Müller ME (1958). A note on the generation of random normal deviates. *Ann Math Stat* **28**: 610–611.
- Chen F, Drezner Z, Ryan J and Simchi-Levi D (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Mngt Sci* **46**: 436–443.
- Chen F, Ryan J and Simchi-Levi D (2000). The impact of exponential smoothing forecasts on the bullwhip effect. *Nav Res Log* **47**: 271–286.
- De Bodt MA and Van Wassenhove L (1983). Cost increases due to demand uncertainty in MRP lot sizing. *Decis Sci* **14**: 345–361.
- De Bodt MA and Van Wassenhove L (1983). Lot sizing and safety stocks in MRP: A case study. *Production and Inventory Mngt* **24**(1): 1–16.
- Dolgui A and Prodron C (2007). Supply planning under uncertainties in MRP environments: A state of the art. *Annu Rev Control* **31**: 269–279.
- Enns ST (2002). MRP performance effects due to forecast bias and demand uncertainty. *Eur J Opl Res* **138**: 87–102.
- Fildes R and Beard C (1992). Forecasting systems for production and inventory control. *Int J Ops Prod Mngt* **12**(5): 4–27.
- Fildes R and Goodwin P (2007). Against your better judgment: Do organizations follow best practice when applying management judgement to forecasting? *Interfaces* **37**: 570–576.
- Fildes R and Kingsman B (1997). *Demand uncertainty in MRP systems: The value of forecasting*. Department of Management Science Working Paper 01/97, Lancaster University.
- Fildes R, Hibon M, Makridakis S and Meade N (1998). Generalising about univariate forecasting methods: Further empirical evidence. *Int J Forecasting* **14**: 339–358.
- Fildes R, Goodwin P, Lawrence M and Nikolopoulos K (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *Int J Forecasting* **25**: 3–23.
- Forrester JW (1961). *Industrial Dynamics*. MIT Press: Cambridge, Mass.
- Gardner Jr ES (1990). Evaluating forecast performance in an inventory control system. *Mngt Sci* **36**: 490–499.
- Jeunet J (2006). Demand forecast accuracy and performance of inventory policies under multi-level rolling schedule environments. *Int J Prod Econ* **103**: 401–419.
- Johnston FR, Taylor SJ and Oliveria RMMC (1988). Setting company stock levels. *J Opl Res Soc* **39**: 15–21.
- Law AM and Kelton WD (2000). *Simulation Modelling and Analysis*. McGraw Hill: New York.
- Lee HL, Padmanabhan V and Whang S (1997). The bullwhip effect in supply chains. *Sloan Mngt Review* **38**(3): 93–102.
- Lee H, So KC and Tang CS (2000). The value of information sharing in a two-level supply chain. *Mgmt Sci* **46**: 626–643.
- Lee TS and Adam Jr EE (1986). Forecasting error evaluation in material requirements planning (MRP) production-inventory systems. *Mngt Sci* **32**: 1186–1205.
- Lee TS, Adam EE and Ebert RJ (1987). An evaluation of forecast error in master production scheduling for material requirements planning systems. *Decis Sci* **18**: 292–307.
- Lee TS, Cooper FW and Adam Jr EE (1993). The effects of forecasting errors on the total cost of operations. *Omega* **21**: 541–550.
- Makridakis S and Hibon M (2000). The M-3 competition: results, conclusions and implications. *Int J Forecasting* **16**: 451–476.
- Newbold P and Bos T (1989). On exponential smoothing and the assumption of deterministic trend plus white noise data-generating models. *Int J Forecasting* **5**: 523–527.
- Orlicky J (1975). *Material Requirements Planning*. McGraw-Hill: New York.
- Ritzman LP and King BE (1993). The relative significance of forecast errors in multistage manufacturing. *J Ops Mngt* **11**: 51–65.
- Silver EA, Pyke DF and Peterson R (1998). *Inventory Management and Production Planning and Scheduling*, 3rd edn. Wiley: New York.
- Vargas GA and Dear RG (1991). Managing uncertainty in multilevel manufacturing systems. *Integr Manufac Syst* **2**(4): 14–26.
- Voss S and Woodruff DL (2006). *Introduction to Computational Optimization Models for Production Planning in a Supply Chain*. Springer: Berlin.
- Wemmerlöv U (1986). A time-phased order-point system in environments with and without demand uncertainty: A comparative analysis of non-monetary performance variables. *Int J Prod Res* **24**: 343–358.
- Wemmerlöv U (1989). The behavior of lot-sizing procedures in the presence of forecast errors. *J Ops Mngt* **8**(1): 37–46.

- Wemmerlöv U and Whybark DC (1984). Lot-sizing under uncertainty in a rolling schedule environment. *Int J Prod Res* **22**: 467–484.
- Xie J, Lee TS and Zhao X (2004). Impact of forecasting error on the performance of capacitated multi-item production systems. *Comput Ind Eng* **46**: 205–219.
- Xu K, Dong Y and Evers PT (2001). Towards better coordination of the supply chain. *Transport Res Part E: Logist Transport Rev* **37**(1): 35–54.
- Zhao X, Xie J and Lau RSM (2001). Improving the supply chain performance: Use of forecasting models versus early order commitments. *Int J Prod Res* **39**: 3923–3929.
- Zhao X, Xie J and Leung J (2002). The impact of forecasting model selection on the value of information sharing in a supply chain. *Eur J Opl Res* **142**: 321–344.
- Ziliak ST and McCloskey DN (2004). Size matters: The standard error of regressions in the American Economic Review. *J Socio Econ* **33**: 527–675.

*Received March 2009;
accepted March 2010 after four revisions*