Proceedings of the 2002 Winter Simulation Conference
E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds.

USE OF GI/G/1 QUEUING APPROXIMATIONS TO SET TACTICAL PARAMETERS FOR THE SIMULATION OF MRP SYSTEMS

S. T. Enns

Dept. of Mechanical & Manufacturing Engineering
University of Calgary
Calgary, AB, T2N 1N4, CANADA

Sangjin Choi

Dept. of Technology Transfer
Korea Institute of Energy Research
Yusong, Taejon, 305-343, SOUTH KOREA

ABSTRACT

There is a lack of prescriptive methods for setting lot sizes and planned lead times effectively in MRP systems. Recent research has suggested the application of queuing relationships. This study experimentally investigates the use of GI/G/1 relationships for lot size selection along with the use of exponentially smoothed feedback for dynamic planned lead time setting. Results show that assumptions regarding lot interarrival time variability have a large effect on lot sizes and performance.

1 INTRODUCTION

Material Requirements Planning (MRP) logic is pervasive in manufacturing planning systems for batch production environments. However, there is currently a lack of good prescriptive methods for setting various tactical parameters, such as lot sizes and planned lead times. It is well recognized that these can have a very significant effect on performance. In fact, much of the criticism related to MRP, such as high work-in-process inventory and long lead times, results from the improper setting of tactical inputs.

The study of work flow in production systems is often undertaken using discrete-event simulation. Simulation is especially appropriate where the production environment is stochastic, exhibits non-stationary loading patterns, and relies on complex planning and control systems. These characteristics apply to typical batch production systems using MRP. In particular, the need to model both material and information flow makes the use of simulation attractive and perhaps necessary.

However, simulation is an experimental approach that can provide little direct guidance as to appropriate settings for MRP tactical parameters. Although extensive experimentation is possible to find good combinations of settings in a very small system, this is not feasible for real world systems. A better alternative is to discover analytical relationships that provide insight and guidance. Recently there has been interest in using queuing relationships to develop methods for lot size selection and planned lead time setting. This approach relies on developing appropriate relationships for multi-item capacity-constrained systems where the entities are lots of parts. An important objective is to then determine optimal lot sizes for all item types, based on lot flowtime (and work-in-process inventory) minimization. Furthermore, the predicted lot flowtimes may guide in setting planned lead times. Some success has been achieved at developing relationships for single resource systems. However, the problem becomes much more complicated when there are networks of resources and when part commonality, hierarchical coordination, assembly operations and time-phased order releases come into play. Investigation is required to determine whether relatively simple models based on queuing relationships can be effectively applied in complex production environments using MRP logic.

A number of studies have provided foundations for exploratory work with MRP systems. Jonsson and Silver (1985) demonstrated that inventory in queue is an important component of costs in capacity-constrained systems, a fact that most discussions on MRP lot sizing tend to ignore. Lambrecht and Vandaele (1996) developed a search procedure for determining optimal lot sizes for the multi-item, single resource problem under GI/G/1 queuing assumptions. Lambrecht, Iven and Vandaele (1998) extended the investigation to look at the multi-item, multi-resource problem in a job shop context. Hill and Raturi (1992) developed an approach, based on M/G/c queuing assumptions, to set the reorder intervals for the POQ lot sizing policy. Their study suggested the queuing relationships could also be used in planned lead time setting but did not experimentally test the performance.

This study examines MRP performance effects when lot sizes are based on minimizing lot flowtimes under GI/G/1 queuing assumptions. One of the main difficulties is that non-stationary arrival patterns and part coordination issues make it difficult to determine appropriate lot interarrival time distributions in an MRP context. Delays due to the time-phased nature of releases further complicate flowtime prediction. In this research the interarrival time coefficients of variation are assumed. The objective is to determine how these affect performance and whether the use of M/G/1 queuing assumptions, which are much easier to deal with, provide lot sizes that lead to satisfactory performance.

As well, performance is compared using dynamically set planned lead times based on order flowtime feedback and shop flowtime feedback.

2 EXPERIMENTAL METHODOLOGY

This section briefly describes the simulation test bed, the scenarios for the production and planning systems, the lot size selection process, the planned lead time setting mechanisms, and the experimental design.

2.1 Experimental Test Bed

An experimental test bed to explore MRP performance requires two modules; a planning system and a simulation program to emulate production activity. A test bed especially designed to be simple, flexible and transparent was used in this research. The planning module was developed within an Excel workbook. Worksheets were used for the user interface, while the MRP logic was implemented through extensive use of Visual Basic for Application (VBA) macros. The production simulation module was implemented in ARENA 5.0 software (Kelton et. al, 2002). Communication between the two modules was facilitated using VBA.

Figure 1 provides an illustration of the test bed components. The solid arrows represent information flow while the dashed arrows represent material flow. The lower section of the diagram represents the shop being simulated. Batches, or lots, of parts are represented as circles while resources are represented as squares. Further details of the test bed may be found in Enns (2001 or 2002a).

2.2 Production Scenario

The material plan was regenerated once per period, with the period assumed to be one week. The MRP system was operated as a bucketed system with 20 time buckets per period (bpp=20). Assuming 5 working days per period, this is the equivalent of 4 time buckets per day. One fifth of the period requirements, as determined by the MPS, was assumed to be required at the start of each day. The shipping policy was such that one fifth of actual demand was shipped at the start of each day. Therefore, the timing of MPS planned requirements and demand shipments coincided.

Two finished goods (independent-demand) products, P1 and P2, were assumed. The demand patterns were assumed to be sinusoidal, with a cycle length of 52 periods (one year). The mean demand was assumed to be 1500 units per period and the amplitude of the demand pattern was assumed to be 360 units, for both products. However, the demand patterns for the two products were offset 26 periods from each other. This allowed a reasonably stable shop load to be maintained, even though the product mix varied considerably through time. No forecast error was assumed since it was desirable not to confound the observed behavior by introducing effects due to demand uncertainty. Figure 2 illustrates the demand patterns used.

The product structures for the two finished goods products are shown in Figure 3. This figure also provides information on the lot setup times and part processing times for all part types.

The shop was assumed to have four machines, with each machine processing two part types. Routing information is also shown in Figure 3. Each part type was assumed

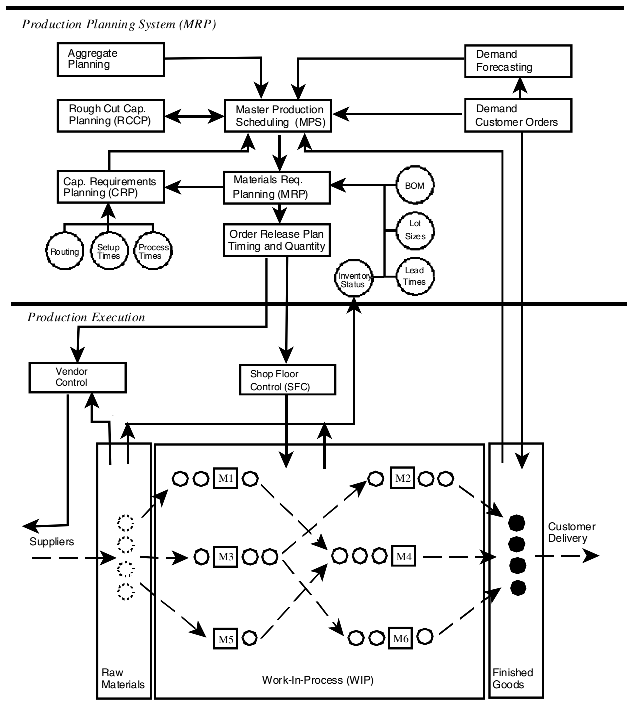

Figure 1: MRP Test Bed

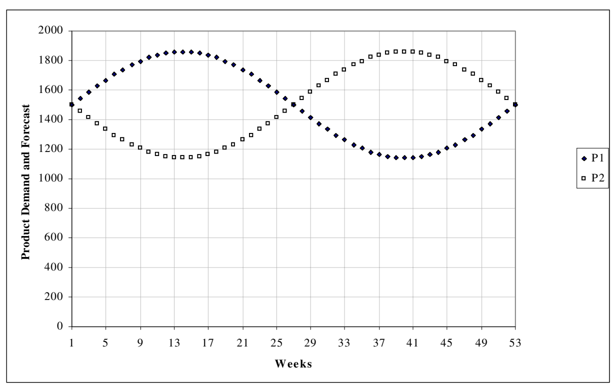

Figure 2: Finished Goods Demand Patterns

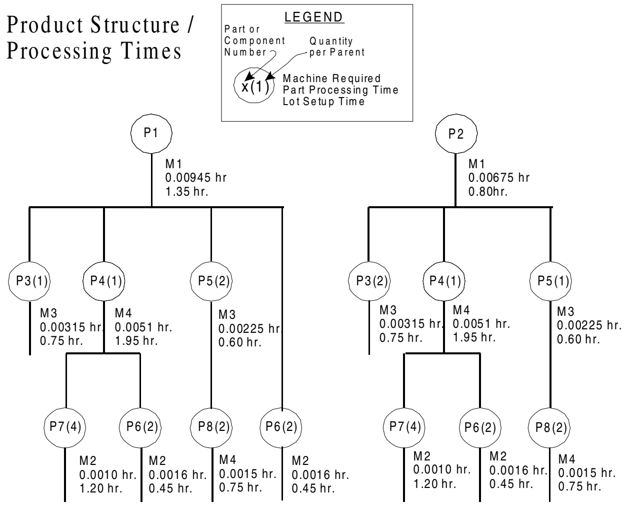

Figure 3: Product Structure

to have only one processing stage. An average scrap rate of three percent, with a negative exponential distribution, was assumed at each processing stage.

2.3 Lot Size Setting

The Fixed Order Quantity (FOQ) lot sizing policy was used for all parts. The size of lots was based on minimizing the weighted average lot flowtimes at each machine m, W_m, under GI/G/1 queuing assumptions. The following simple approximation was used to estimate the weighted average flowtimes (Whitt, 1983). This relationship is exact under M/G/1 queuing assumptions, where Poisson arrivals mean the variation of lot interarrival times, $c_{a,m}$, is 1.

$$W_m = W_{q,m} + \bar{x}_m = \bar{x}_m \frac{\left(c_{a,m}^2 + c_{s,m}^2\right)}{2} \frac{\rho_m}{1-\rho_m} + \bar{x}_m \tag{1}$$

where $W_{q,m}$ is the weighted mean time in queue, $\bar{x}_m$ is the weighted mean lot service time, $c_{a,m}$ is the coefficient of variation for lot interarrival times, $c_{s,m}$ is the coefficient of variation for lot service times and ρ_m is the machine utilization rate.

The weighted mean lot service time, including setup times, for n part types processed on machine m is given by the following,

$$\bar{x}_m = \frac{\sum_{j=1}^{n} \frac{\bar{D}_j}{Q_j}\left[\tau_j + \frac{Q_j}{P_j}\right]}{\sum_{j=1}^{n} \frac{\bar{D}_j}{Q_j}} \tag{2}$$

where j is the part type index, $\bar{D}_j$ is the average demand rate, Q_j is the part type lot size, P_j is the part processing rate and τ_j is the lot set up time. $\bar{D}_j$ is based on average demand for independent demand products and average requirements for dependent demand parts, adjusted for average scrap rates.

The utilization rate, including setup times, on machine m is then given by the following.

$$\rho_m = \sum_{j=1}^{n} \left[\frac{\bar{D}_j}{Q_j}\left(\tau_j + \frac{Q_j}{P_j}\right)\right] \tag{3}$$

If it is assumed the lot setup times and part processing times are deterministic, the coefficient of variation of lot service times squared is expressed as follows,

$$c_{s,m}^2 = \frac{\sum_{j=1}^{n} \frac{\bar{D}_j}{Q_j}\left[\tau_j + \frac{Q_j}{P_j}\right]^2 \left(\sum_{j=1}^{n} \frac{\bar{D}_j}{Q_j}\right)^{-1}}{\bar{x}_m^2} - 1 \tag{4}$$

Following substitution, the wieghted mean lot flowtime at machine m can be written as follows,

$$\begin{aligned} W_m &= W_{q,m} + \bar{x}_m \\ &= \frac{\bar{x}_m^2 \sum_{i=1}^{n} \frac{\bar{D}_j}{Q_j}\left(c_{a,m}^2 - 1\right) + \sum_{i=1}^{n} \frac{\bar{D}_j}{Q_j}\left(\tau_j + \frac{Q_j}{P_j}\right)^2}{2\left(1-\rho_m\right)} + \bar{x}_m \end{aligned} \tag{5}$$

The GI/G/1 approximations, given by Equations (1) to (5), allow average flowtimes to be estimated (Whitt, 1983). To find lot sizes that minimize average flowtimes, we need to find the partial derivative of W_m with respect to the lot size for each part type j processed on machine m. The resulting set of simultaneous equations can then be set equal to zero and solved to determine optimal lot sizes. The differential equations to minimize lot flowtimes as a function of lot sizes for each part type were derived as follows,

$$\frac{\partial W_m}{\partial Q_j} = AA + BB + CC + DD = 0 \tag{6}$$

where

$$AA = \frac{\left\{\left[\frac{2\rho_m\left(\frac{-\bar{D}_j\tau_j}{Q_j^2}\right)\sum_{j=1}^{n}\frac{\bar{D}_j}{Q_j} + \rho_m^2\frac{\bar{D}_j}{Q_j^2}}{\left(\sum_{j=1}^{n}\frac{\bar{D}_j}{Q_j}\right)^2}\right]\left(c_{a,m}^2 - 1\right)\right\}2\left(1-\rho_m\right)}{\left[2\left(1-\rho_m\right)\right]^2},$$

$$BB = \frac{\left[\frac{-\overline{D}_j}{Q_j^2}\left(\tau_j + \frac{Q_j}{P_j}\right)^2 + 2\frac{\overline{D}_j}{Q_j P_j}\left(\tau_j + \frac{Q_j}{P_j}\right)\right] 2\left(1-\rho_m\right)}{\left[2\left(1-\rho_m\right)\right]^2},$$

$$CC = -\frac{\left\{\left[\left[\frac{\rho_m{}^2}{\left(\sum_{j=1}^{n}\frac{\overline{D}_j}{Q_j}\right)}\left(c_{a,m}^2 - 1\right) + \sum_{j=1}^{n}\frac{\overline{D}_j}{Q_j}\left(\tau_j + \frac{Q_j}{P_j}\right)^2\right] 2\left(\frac{\overline{D}_j \tau_j}{Q_j^2}\right)\right]\right\}}{\left[2\left(1-\rho_m\right)\right]^2},$$

$$DD = \frac{\left[\frac{\overline{D}_j}{Q_j P_j} - \frac{\overline{D}_j}{Q_j^2}\left(\tau_j + \frac{Q_j}{P_j}\right)\right]\sum_{j=1}^{n}\frac{\overline{D}_j}{Q_j} + \rho_m \frac{\overline{D}_j}{Q_j^2}}{\left(\sum_{j=1}^{n}\frac{\overline{D}_j}{Q_j}\right)^2}.$$

The resulting set of j simultaneous equations were used to solve for the optimal lot sizes of all parts processed on machine m. The Excel Solver add-in was used for this purpose.

There was assumed to be no uncertainty in setup and processing times and machines were assumed not to break down. Lots in queue were processed on the basis of earliest part due date. No lot splitting was allowed.

2.4 Planned Lead Time Setting

The planned lead times were set dynamically based on exponentially smoothed feedback. The two feedback alternatives were to base planned lead times on the order flowtimes or on the shop floor flowtimes. The order flowtimes are defined to be the time between an internal order release, generated by the MRP system, and the time the order is completed. The shop floor flowtimes are defined to be the time between when all component parts required for a given internal order are ready and the time the order is completed. If the component parts are available at the time an order is released, the order flowtime and the shop flowtimes will be equal. However, if at least one component part is not available to meet the order release requirements, the order flowtime will be longer than the shop flowtimes. The assumption was made that component parts entered the queue for the next stage of processing only when all components were available in sufficient quantities to complete the lot for the released order.

The relationships used in dynamic planned lead time setting can be clarified by referring to Figure 4. Further analysis of exponential smoothing applied to dynamic planned lead time setting can be found in Enns (2002b).

The order, $F(R)_{i,j}$, and shop, $F(A)_{i,j}$, flowtimes for lot i of part type j are determined as follows,

$$F(R)_{i,j} = C_{i,j} - R_{i,j} \quad (7)$$

$$F(A)_{i,j} = C_{i,j} - A_{i,j} \quad , where\ A_{i,j} \geq R_{i,j} \quad (8)$$

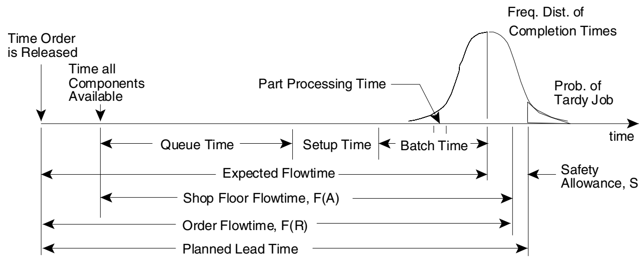

Figure 4: Flowtime and Planned Lead Time references

where $C_{i,j}$ is the completion time for the lot, $R_{i,j}$ is the order release time and $A_{i,j}$ is the time all components are available to process the order.

The exponentially smoothed estimate of future order and shop flowtimes for each part type j can then be determined as follows,

$$\hat{F}(R)_{i+1,j} = \hat{F}(R)_{i,j}(1-\alpha) + F(R)_{i,j}\alpha \quad (9)$$

$$\hat{F}(A)_{i+1,j} = \hat{F}(A)_{i,j}(1-\alpha) + F(A)_{i,j}\alpha \quad (10)$$

where α is the smoothing constant.

The planned lead time for part type j at any given time t, $PLT_{t,j}$, can then be determined for the two methods respectively as follows,

$$PLT_{t,j} = PLT_{t,j} + \frac{1}{bpp}\left\lceil\left(\hat{F}(R)_{i+1,j} - \left(PLT_{t,j} - S_j\right)\right)bpp - 0.5\right\rceil \quad (11)$$

$$PLT_{t,j} = PLT_{t,j} + \frac{1}{bpp}\left\lceil\left(\hat{F}(A)_{i+1,j} - \left(PLT_{t,j} - S_j\right)\right)bpp - 0.5\right\rceil \quad (12)$$

where bpp is the number of buckets per period and S_j is a safety flow allowance for part type j, equal the time interval for an integer number of buckets. The planned lead times are rounded to the nearest time bucket by subtracting 0.5 and then rounding up the the nearst integer, as designated by the $\lceil\ \rceil$ brackets. These equations assume that the exponentially smoothed flowtimes, planned lead times and safety allowances are specified in terms of periods.

2.5 Experimental Design

The experimental design consisted of three factors. The first factor was the coefficient of variation of lot interarrival times at each machine, $c_{a,m}$, used to determine lot sizes for each part type. These values were set at levels 0.25, 0.50, 0.75 and 1.00. The values of $c_{a,m}$ were assumed to be equal for each of the four machines at any given combination of experimental settings. The optimal lot sizes for part types P1 to P8, using Equation (6) with each $c_{a,m}$ value, are shown in Table 1. The shop utilization rates ranged from 91.5% when lot sizes were based on a $c_{a,m}$ of 0.25 to 84% when lot sizes were based on a $c_{a,m}$ of 1.00.

Table 1: Lot Sizes for various $c_{a,m}$ values

	Lot Sizes			
	$c_{a,m}$ =0.25	$c_{a,m}$ =0.50	$c_{a,m}$ =0.75	$c_{a,m}$ =1.00
P1	266	345	420	497
P2	401	438	442	423
P3	536	674	791	891
P4	1316	1690	2025	2379
P5	773	907	988	1020
P6	1892	1974	1879	1713
P7	2789	3703	4615	5586
P8	4708	5039	4792	4255

The second factor was the method of dynamic planned lead time setting. This factor was run at two levels. The first used order flowtime feedback, as given by Equation (11). The second used shop flowtime feeback, as given by Equation (12). The smoothing constant, α, was set at 0.05 in both cases.

The third factor was the safety flow allowance setting, S_j, used to control delivery performance for finished goods. This factor was set at eight levels for both P1 and P2. These levels ranged from 0 to 0.35 time periods, in increments of 0.05 periods. This increment corresponded to the length of one time bucket in the MRP planning system. The value of S_j for all dependent demand parts was set to zero.

The 64 combinations of experimental settings (4 levels of lot interarrival time variability, 2 planned lead time setting approaches, and 8 safety factor settings) were each run for three replications. These replications were 260 periods long (5 years). The first 52 periods were used for initialization, while the remainder of the run was used for data collection. Common random numbers were used for the distribution parameters used to determine scrap quantities. Within-group variance was low.

Performance analysis was based on inventory and customer delivery measures. The primary inventory measure was based on processing time invested. This measure is preferable to inventory counts because it weights inventory according to the amount of value-added activity completed. The customer delivery performance was measured in terms of both delivery mean tardiness and the proportion of customer orders delivered immediately from stock.

3 DISCUSSION OF RESULTS

Figure 5 shows the delivery mean tardiness versus inventory investment results when order flowtime feedback was used to set planned lead times. The individual lines represent performance when $c_{a,m}$ values of 0.25, 0.50, 0.75 and 1.00 were used to set lot sizes. The points moving right along the curve correspond to increasing values of S_j. Figure 6 is similar to Figure 5 except that the proportion of deliveries on time is plotted versus inventory investment.

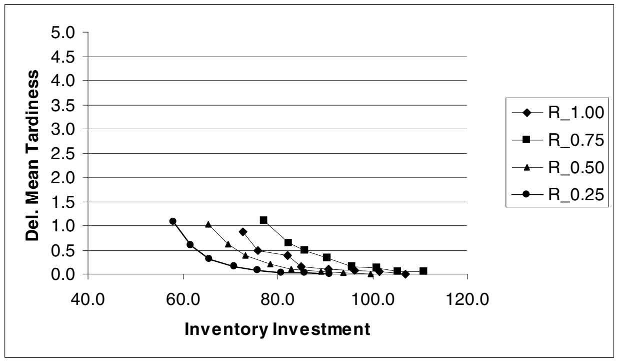

Figure 5: Delivery Mean Tardiness vs Inventory

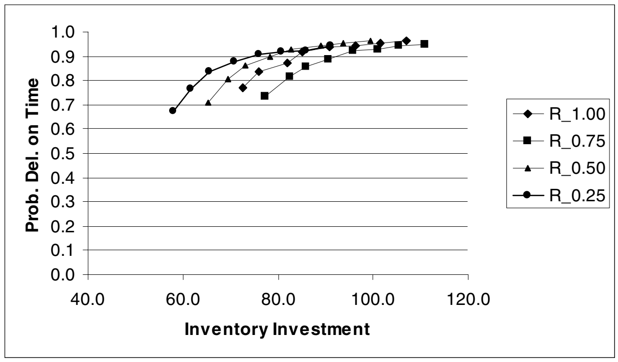

Figure 6: On Time Delivery vs Inventory

The results showed improved performance as the $c_{a,m}$ values used to determine lot sizes were reduced. Figure 5 shows that delivery mean tardiness was substantially lower for a given level of inventory investment when lot sizes were based on a $c_{a,m}$ of 0.25 versus 1.0. Likewise, the proportion of deliveries on time, shown in Figure 6, increased for a given level of inventory investment when lot sizes were based on a $c_{a,m}$ of 0.25 versus 1.0. When $c_{a,m}$ values were decreased below 0.25 (not shown), performance deteriorated and became erratic. It appeared this was due to queues building up as a result of excessively small lot sizes and correspondingly high utilization levels. Therefore, it seems that using a $c_{a,m}$ value somewhere around 0.25 for determining lot sizes resulted in the best performance for the production scenario tested.

The results obtained using shop flowtime feedback to set planned lead times followed very similar patterns. However, the performance was not quite as good, especially when lot sizes were based on high $c_{a,m}$ values. Figure 7 shows a comparison of the delivery mean tardiness versus inventory investment results when the two types of feedback were used for planned lead time setting. For clarity, only the results when $c_{a,m}$ values of 0.25 and 1.00 were assumed are shown. The line marked A_1.00 designates that a $c_{a,m}$ value of 1.00 was used along with planned lead times based on shop flowtime feedback, while R_1.00 designates that a $c_{a,m}$ value of 1.00 was used along with planned lead times based on order flowtime feedback. Similarly, A_0.25 and R_0.25 indicate a $c_{a,m}$ of 0.25 was assumed. Figure 8 shows similar results for the proportion of deliveries on time versus inventory investment.

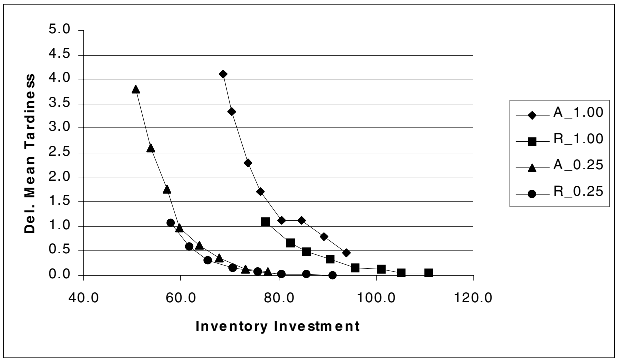

Figure 7: Delivery Mean Tardiness Comparison

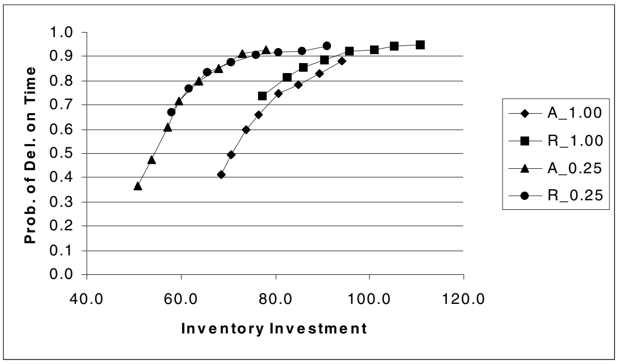

Figure 8: On Time Delivery Comparison

From these results it can be observed that dynamic planned lead times set on the basis of order flowtimes resulted in better performance than those set on the basis of shop flowtimes. Therefore, it is better to monitor the time it takes to complete an order released to the shop floor rather than the time it takes to process the lot once all the components are available. However, Figures 7 and 8 show that the differences between performance using planned lead time settings based on order flowtime versus shop flowtime feedback decreased as the $c_{a,m}$ values used for lot sizing decreased. This suggests that choosing the method of feedback used to dynamically set planned lead times is less important when lot sizes are properly set. In other words, appropriate lot size selection means there will be less difference between order flowtimes and shop flowtimes.

4 CONCLUSIONS

This research has shown that assumptions regarding lot interarrival time variability make a big difference when queuing relationships are used for lot sizing. Significant differences in lot sizes affect both inventory and delivery performance. It further appears that the assumption of Poisson arrivals and use of M/G/1 queuing relationships may lead to excessively large lot sizes. When lot sizing is based on properly selected interarrival time variability, dynamic planned lead time adjustment based on either order flowtime feedback or shop flowtime feedback would appear to work well. Further research is required to measure and select $c_{a,m}$ values that can be used effectively in MRP lot sizing.

ACKNOWLEDGMENTS

This research was supported in part by a grant from the National Science and Engineering Research Council (NSERC) of Canada.

REFERENCES

Enns, S.T. 2001. MRP performance effects due to lot size and planned lead time settings, *International Journal of Production Research*, Vol. 39, No. 3, 461-480.

Enns, S.T. 2002a. MRP performance effects due to forecasting bias and demand uncertainty, *European Journal of Operational Research*, Vol. 138, No. 1, 87-102.

Enns, S.T. 2002b. Work load responsive adjustment of MRP planned lead times, In *Proceedings of the 2nd International Responsive Manufacturing Conference*, ed. A. Baykasoglu and T. Dereli, 658-662. Gaziantep, Turkey.

Hill, A.V., and A.S. Raturi. 1992. A model for determining tactical parameters for Materials Requirements Planning systems, *Journal of the Operational Research Society*, Vol. 43, 605-620.

Jonsson, H. and E.A. Silver. 1985. Impact of processing and queueing times on order quantities, *Material Flow*, Vol. 2, 221-230.

Kelton, W.D., R.P. Sadowski and D.A. Sadowski. 2002. *Simulation with Arena*, 2nd Ed., McGraw-Hill Co.

Lambrecht, M.R., and N.J. Vandaele. 1996. A general approximation for the single product lot sizing model with queueing delays, *European Journal of Operational Research*, Vol. 95, 73-88.

Lambrecht, M.R., P.L. Ivens, and N.J. Vandaele. 1998. ACLIPS: A Capacity and Lead Time Integrated Procedure for Scheduling, *Management Science*, Vol. 44, 1548-1561.

Whitt, W. 1983. The Queueing Network Analyzer, *The Bell Systems Technical Journal*, Vol. 62, No. 9, 2779-2813.

AUTHOR BIOGRAPHIES

SILVANUS T. ENNS is an Associate Professor in the Department of Mechanical and Manufacturing at the University of Calgary, Calgary, Canada. He received his PhD in Operations Management from the University of Minnesota. His main research interest lies in the development of algorithms to support enhanced MRP performance. As well, he has worked on various aspects of job shop, batch production and supply chain modeling and analysis. His email address is <enns@ucalgary.ca>.

SANGJIN CHOI is an Associate Researcher in the Technology Transfer Team within the Korea Institute of Energy Research. He graduated in Industrial Engineering from Pusan National University, Pusan, Korea in 2000 and then worked as a post-doctoral Research Associate in the Department of Mechanical and Manufacturing at the University of Calgary until 2001. His research interests lie in the areas of decision making within stochastic inventory and production management environments.