

## **INVESTIGATING THE EFFECT OF DEMAND AGGREGATION ON THE PERFORMANCE OF AN (R, Q) INVENTORY CONTROL POLICY**

Manuel D. Rossetti  
Mohammad Shbool  
Vijith Varghese  
Edward Pohl

University of Arkansas  
Department of Industrial Engineering  
4207 Bell Engineering Center  
Fayetteville, AR 72701, USA

### **ABSTRACT**

This paper investigates the effect of demand aggregation on the performance measures of an inventory system controlled by a (r, Q) policy. Demand usage data is available at different time scales, i.e., daily, weekly, monthly etc., and forecasting is based on these time scales. Using forecasts, appropriate lead time demand models are constructed and used in optimization procedures. The question being investigated is what effect the forecasting time bucket has on whether or not the inventory control model meets planned performance. A simulation model is used to compare performance under different demand aggregation levels. The simulation model of the optimized (r, Q) inventory system is run for the planning horizon and the supply chain operational performance measures like ready rate, expected back order etc., are collected. Subsequently, the effect of aggregating the demand and planning accordingly is analyzed based on the simulated supply chain's operational performance.

### **1 INTRODUCTION**

Inventory control involves complex trade-offs between costs and services. A key to the successful execution of inventory control systems is setting the policy parameters so as to minimize cost subject to constraints on service. Once the policy parameters are set, it is expected that the planned performance of the system will be met; however, often this is not the case. When customer service measures are not met, not only do customers lose, but inventory managers begin to question whether or not the models are worth the effort. This paper investigates one of the reasons that planned performance might not be achieved: the process by which a forecast is translated to appropriate policy parameters. In particular, we investigate the effect that different aggregation levels have on a supply chain's operational performance.

The time bucket size forms the basic planning period for managing the system's behavior over time. We might have a time bucket size of 1 day. In that case, we plan for daily demand over a horizon of so many days. Other common time bucket sizes are weekly, monthly, quarterly, etc. The time bucket size often forms the basis for the forecasting period. For example, if the total demand for a stock-keeping-unit (SKU) is collected daily, then the forecasting period might be daily. Alternatively, we might aggregate daily demand into weekly demand, and make forecasts based on a weekly forecast period. The basic purpose of this paper is to investigate via simulation the effect that the size of the forecasting period has on the ability of the inventory models to meet their planned performance. When observed demand from smaller time bucket sizes are totaled into larger time bucket sizes, this is called temporal aggregation.

The effect of temporal aggregation is complicated by the processes used to translate a forecast into optimal policy parameters. In this paper, simulation is used to understand these effects.

The remainder of this paper is structured as follows. Section 2 provides a literature review of some work done in the context of intermittent demand and inventory control performance under temporally aggregated demand. Section 3 (Research Methodology) presents a summary of the underlying concepts (inventory modeling, forecast modeling, lead time demand modeling, temporal aggregation, and measuring effects of temporal aggregation). Section 4 describes the simulation model, the experiments, and the results. The paper concludes with some ideas for future work.

## **2 LITERATURE REVIEW**

Forecasting is the first step when developing inventory control policies. Krajewski (2010) mentions three key points concerning forecasting. The first is that the forecast will never be without error due to the stochastic nature of demand. Secondly, forecasting for groups is more accurate than it is for individuals. This point touches directly on the core research question within this paper, namely how demand is aggregated. Lastly, Krajewski (2010) mentions the commonly accepted notion that as the time horizon increases, the forecasting accuracy decreases. Thus, the forecast directly effects inventory control.

The impact of forecasting on inventory control was studied by Gardner (1990) in a large physical distribution system. This is important since the research investigated the effect of the forecasting technique on the inventory policy performance. Gardner showed that choosing the forecasting technique would make a difference in determining inventory investment and customer service. His methodology was to compare the forecasting techniques in the form of tradeoff curves between inventory investment and customer service. More than just comparing forecasting techniques, Sani and Kingsman (1997) provided a comparison of different periodic inventory policies in addition to comparing forecasting methods, and investigated which are more appropriate for demands with low and intermittent patterns. They claimed that (s,S) heuristic procedures, Naddor's heuristic, the power and normal approximations are the most recommended as the best periodic inventory policies for the mentioned demand patterns. Moreover, 52 week moving average was found to be the best forecasting technique, followed by Croston's method. The comparison of ten inventory policies with five forecasting methods tested for each was carried out using simulation and statistical analysis. The comparison was based on two performance measures, annual inventory costs including the shortage penalty costs and the customer service level (Ready Rate). The methodology was basically to test if there is a significant difference between the results of the two performance measure produced by the ten inventory policies when combined with each of the five forecasting methods. A nonparametric method (Friedman's test) was used instead of ANOVA for the analysis.

Developing a good forecast depends on the characteristics of the demand process. Classification of demand according to its characteristics can provide insights into which forecasting techniques to utilize. This paper focuses on demand that is considered intermittent in nature. According to Nikolopoulos et al. (2010), "intermittent demand patterns are characterized by infrequent demand arrivals coupled with variable demand sizes." The existence of many zero demand events in the data series is a major problem and adds difficulty to the forecasting process. Babai et al. (2012) indicate that the use of traditional forecasting techniques are often ineffective when demand is intermittent or lumpy in nature. Zero demands, being the main problem, can be reduced by aggregating demand within specific time buckets thereby producing a lower-frequency demand pattern with fewer zero demands. This aggregation strategy is referred to as temporal aggregation within the literature. However, when the demand is aggregated in this manner, there are less data points in the series and the characteristics of the stochastic process will be changed. For example one drawback in aggregating demand is the loss of information since the number and frequency of data points is reduced and their temporal location is lost.

Eaves and Kingsman (2004) used weekly, monthly, and quarterly aggregation levels in their research. They worked to assess the performance of forecasting techniques used in the estimation of intermittent or slow-moving demand. Forecasting methods considered for comparison were an approximation method

created by Syntetos and Boylan (2006), exponentially smoothing, Croston's method, a 12-month moving average and a simple previous year average method. They found that the approximation method was the best forecasting method when controlling spare parts inventory. They also proposed a measure of accuracy instead of traditional measures for forecasting techniques in the case of slow-moving and intermittent demand, namely implied stock holding. It is defined as the safety margin that results in a stock out quantity of zero. Syntetos and Boylan (2006) provide an assessment of the empirical performance of inventory control under intermittent demand estimation procedures. Four forecasting methods were considered in the study, simple moving average, single exponential smoothing, Croston's method, and a new method which they developed. The comparison was based on two performance measures, customer service level and stock volume differences. They showed that the newly developed forecasting technique was the best overall. This research is unique in that rather than focusing on forecast error to select the forecasting technique, the authors based the decision on inventory performance driven by the forecast.

Nikolopoulos et al. (2010) investigated the effect of aggregation on forecast accuracy. An aggregate-disaggregate intermittent demand (ADIDA) approach was developed. They investigated an optimum aggregating level empirically, as well as the effect of aggregation when bucketing demand in time periods equal to the lead time plus review period. They found that improvements can be achieved by setting the aggregation level equal to the lead time plus the review period. Babai et al. (2012) investigated empirically the effect of temporal aggregation on inventory control performance of intermittent demand estimators. The method depends on comparing forecasting techniques with demand un-aggregated and aggregated using simulation. The results showed that aggregating demand resulted in higher customer service levels. The motivation for the research in this paper stems partly from the approach presented in Nikolopoulos (2010) and Babai et al. (2012). However, this research is different since it considers a comparison between the analytical and simulated (actual) performance of the system for different time bucketing strategies for the demand. This paper seeks to find the aggregation level that gives a robust inventory policy with minimal error between predicted and actual estimates. That is, the criteria for measuring the robustness of the inventory policy to the forecast time bucket is the difference between the planned operational performance and the achieved operational performance. We measure this difference via simulation.

### 3 RESEARCH METHODOLOGY

#### 3.1 Inventory Modeling

In this section, we describe the approach taken to assess the effect of temporal aggregation on expected operational performance. The key performance measures of interest are the expected ready-rate ( $\overline{RR}$ ), expected amount backordered ( $\overline{B}$ ), expected inventory level ( $\overline{I}$ ), and the expected total cost ( $\overline{TC}$ ). For an (r, Q) inventory system these values can be computed (or approximated) via the following notation (as per Zipkin (2000)):

Let  $D$  be a random variable representing the demand per period with  $E[D] = \mu_D$  and  $var[D] = \sigma_D^2$

Let  $L$  be a constant representing the replenishment lead time in periods

Let  $LTD$  be a random variable representing the demand during lead time  $E[LTD] = L\mu_D$  and  $var[LTD] = L\sigma_D^2$

Let  $f_{LTD}(x)$  be the probability mass function of LTD  $f_{LTD}(x) = P\{LTD = x\}$

Let  $F_{LTD}(x)$  be the cumulative distribution function of LTD,  $P\{LTD \leq x\}$  and  $F_{LTD}^o = 1 - F_{LTD}(x)$

Let  $F_{LTD}^1(x)$  be the first order loss function of LTD. Define  $(x)^+ = \max(x, 0)$ , then  $F_{LTD}^1(x) = E[(X - x)^+]$

Let  $F_{LTD}^2(x)$  be the second order loss function of LTD, then  $F_{LTD}^2(x) = \frac{1}{2}E[(X - x)^+(X - x - 1)^+]$

Define  $k$  to be the fixed ordering cost in \$/order

- $C$  to be the cost per item, \$/unit
- $i$  to be the inventory carrying charge, \$/\$/time
- $h = i * c$  be the holding cost per item, \$/unit/time

- $b$  be the backorder cost, \$/unit/time
- $r$  to be the reorder point, unit
- $Q$  to be the reorder quantity, unit
- $\overline{OF}(Q)$  to be the expected order frequency
- $\overline{I}(r, Q)$  to be the expected on hand inventory
- $\overline{B}(r, Q)$  to be the expected backorders amount
- $\overline{RR}(r, Q)$  to be the expected ready rate

Then we have that

$$\overline{TC}(r, Q) = k\overline{OF}(Q) + h\overline{I}(r, Q) + b\overline{B}(r, Q) \quad (1)$$

Where

$$\overline{OF}(Q) = \frac{\mu_D}{Q} \quad (2)$$

$$\overline{I}(r, Q) = \frac{1}{2}(Q + 1) + r - \mu_{LTD} + \overline{B}(r, Q) \quad (3)$$

$$\overline{B}(r, Q) = \frac{1}{Q}[F_{LTD}^2(r) - F_{LTD}^2(r + Q)] \quad (4)$$

$$\overline{RR}(r, Q) = \frac{1}{Q}[F_{LTD}^1(r) - F_{LTD}^1(r + Q)] \quad (5)$$

### 3.2 Forecast Modeling

Let us consider how these formulas are operationalized based on a forecast. In practice, a demand series  $\{D_1, D_2, D_3, \dots\}$  will be available, where  $D_i$  is the demand in period  $i$ . Here a period can represent some common time scale e.g. (day, week, month, etc.). In this paper, we assume that the demand is stationary. Analyzing the demand characteristics of the items, the inventory system considered in this research manages primarily, items which are intermittent in nature. Hence, in addition to traditional forecasting approaches Simple Exponential Smoothing and Moving Average we have included intermittent demand forecasting techniques such as MCARTA, Croston and Syntetos approaches. Croston split the intermittent demand time series into two constituent time series, one series for the non-zero demand values and the other for the time interval between consecutive non-zero demands or transactions. He applied exponential smoothing to both of the series. Croston's estimates of mean interval between transaction  $\hat{p}_i$  and Croston's estimates of the mean demand size  $\hat{z}_i$  at time index  $i$  can be easily calculated. Using  $\hat{p}_i$  and  $\hat{z}_i$ , an estimate of the mean demand per period,  $\hat{D}_i$  is made. In this method, the time interval since last non-zero demand  $q$  throughout the forecasting is recorded. At the beginning of a time index  $i$ , if the previous demand given by  $\hat{D}_{i-1}$  is equal to zero, then  $\hat{p}_i = \hat{p}_{i-1}$  and  $\hat{z}_i = \hat{z}_{i-1}$  and  $q$  is updated to  $q+1$ . If  $\hat{D}_{i-1} \neq 0$ , then  $\hat{p}_i$  and  $\hat{z}_i$  are exponentially smoothened as in Equation (5). Finally, a value of 1 is assigned to  $q$  (time since the last non-zero demand). With Croston's estimates  $\hat{p}_i$  and  $\hat{z}_i$ , the mean demand per period,  $\hat{Y}_i$  can be estimated as per Equation (6).

$$\hat{p}_i = \hat{p}_{i-1} + \alpha(q - \hat{p}_{i-1}) \text{ and } \hat{z}_i = \hat{z}_{i-1} + \alpha(\hat{Y}_{i-1} - \hat{z}_{i-1}) \dots (5)$$

$$\hat{Y}_i = \frac{\hat{z}_i}{\hat{p}_i} \dots (6)$$

Syntetos approach updates the  $\hat{p}_i$  and  $\hat{z}_i$  as in Croston's method, but the mean demand per period,  $\hat{Y}_i$  is computed taking into consideration a correction factor  $\lambda_{CF}$  (Equation 7). The approximation method was found to improve the forecasting accuracy in intermittent demand cases when compared to EWMA and Croston's method.

$$Y_i = \left(1 - \frac{\alpha}{2}\right) \frac{\hat{z}_i}{\hat{p}_i} \dots (7)$$

MCARTA is a parametric bootstrapping approach in which the demand occurrence process is modeled as a two stage Markov Chain and the demand amount process is modeled as Auto Regressive to Any model.

Based on the demand characteristics of demand history the demand occurrence ( $X_i$ ) and demand amount processes ( $V_i$ ) are modeled and a bootstrap sample is generated. The mean of the bootstrap sample is the forecast estimate.

Let state at time  $i$ ,  $X_i = 0$  if there is zero demand at time  $i$  and let  $X_i = 1$  if a non-zero demand occurs at time  $i$ . The transition of the process was modeled according to a Markov process where,  $p_{lm} = P\{X_i = m | X_{i-1} = l\}$  represents the transition probabilities. For a bootstrap sample size,  $= 1 \dots B$ , the Markov chain model generates the state values for  $i$ :  $X_i^b = \{X_i^1, X_i^2 \dots X_i^B\} \sim MC(p_{11}, p_{01})$ . If the state value randomly generated is 0, this causes a zero value  $V_i^b = 0$ . Otherwise, a non-zero value is generated according to an assumed underlying non-zero discrete distribution. This distribution is selected based on the discrete distribution fitting procedure in Axsäter (2006). The procedure matches the parameters of the distributions by using the first two moments of non-zero demand. Axsäter's rule selects a discrete demand distribution from one of the distributions of Poisson, negative binomial (also geometric) or binomial distribution. Lag 1 correlation is introduced into the underlying discrete distribution by the Autoregressive to Any Algorithm (ARTA) discussed in Cario and Nelson (1996) and Biller and Nelson (2005). According to Cario and Nelson (1996) and Biller and Nelson (2005), the ARTA considers an autoregressive process, AR(1) which has an underlying non-zero distribution with normal noise  $\sim Normal(0, 1 - \hat{\phi}_{1,NZ}^2)$ . The ARTA generates correlated data  $V_i^b \sim ARTA(G(Y_{NZ}), \phi_{1,NZ})$ , with correlation coefficient,  $\phi_{1,NZ}$  and the underlying non-zero distribution,  $Y_{NZ}$  selected using Axsäter procedure.

### 3.3 Lead Time Demand Modeling

A forecasting technique is used to model the historical demand and develop estimates for the mean forecasted demand  $\hat{\mu}_F$  and the variance of the forecast error  $\sigma_e^2$ . Axsäter (2006) and Silver et al. (1998), suggest that the forecast should be translated into the mean and variance of the lead time demand distribution in the following manner. First the forecast is made for the next period, then assuming that the lead time ( $L$ ) is expressed in the same time units of as the forecast period we have that  $\mu_{LTD} = \hat{\mu}_F * L$ . Let MAD be the mean absolute deviation for the forecast. Then, Nahmias (2009) suggests that this can be converted to the standard deviation of the forecast error via  $\hat{\sigma}_e = 1.25 * MAD$ . The standard deviation of the forecast error is used to specify the variance of the lead time demand. Let  $\sigma_{LTD}$  be the standard deviation of the demand during lead time. Axsäter (2006) and Silver et al. (1998) suggest that  $\sigma_{LTD} = \hat{\sigma}_e L^c$  where  $0.5 \leq c \leq 1$ . Here we choose the common choice of  $c = 0.5$ , so that  $\sigma_{LTD} = \hat{\sigma}_e \sqrt{L}$ . The two parameters,  $(\mu_{LTD}, \sigma_{LTD})$  are then matched to the moments of a suitable distribution. In this paper, we, assume that the demand during lead time is gamma distributed, such that the shape ( $\alpha$ ) and scale ( $\beta$ ) parameters are  $\alpha = \mu_{LTD}^2 / \sigma_{LTD}^2$  and  $\beta = \sigma_{LTD}^2 / \mu_{LTD}$ . This process is discussed in Axsäter (2006), Silver et al. (1998) and Nahmias (2009). See also Silver and Rahnema (1986), Silver and Rahnema (1987), and Strijbosch et al. (1997). While other methods may be used for this process, we used this standard/common procedure. Since all testing will use this same procedure, we expect that the relative comparisons will be valid.

### 3.4 Temporal Aggregation

In order to examine the effect of temporal aggregation on inventory management, we need to define performance indicators that can measure the effect. Let  $\tau$  be the temporal aggregation amount. Suppose that the demand series  $\{D_i\}$  is captured on a daily basis. Then if there is no aggregation by time, then  $\tau = 1$  (daily). For weekly aggregation, we would have  $\tau = 7$ . Suppose we have the following series

$$D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}, D_{14}$$

If we aggregate with  $\tau = 7$ , then in the aggregated series we have

$$D_1(\tau = 7) = D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7$$

$$D_2(\tau = 7) = D_8 + D_9 + D_{10} + D_{11} + D_{12} + D_{13} + D_{14}$$

as the 1<sup>st</sup> two aggregated values. In general, let  $D_j^a(\tau)$  be the aggregate value for the  $j^{th}$  aggregate period

$$D_j^a(\tau) = \sum_{i=1}^{\tau} D_{(j-1)\tau+i} \quad j = 1, 2, \dots, k$$

The  $j^{th}$  period consists of the observations  $D_{(j-1)\tau+1}, D_{(j-1)\tau+2}, \dots, D_{j\tau}$  for  $j=1, 2, \dots, k$  where  $k$  is the total number of aggregate periods. As mentioned in the literature review, a key motivation for constructing an aggregate series  $\{D_j^a(\tau)\}$   $j = 1, 2, \dots, k$  is to affect the properties of the resulting forecast.

Figure 1 illustrates the chain of dependence within the entire inventory management process. Data preparation, forecasting, modeling lead time demand, inventory analysis all lead to planned stocking levels, yielding expected performance. We can now see that aggregation can change which forecasting technique may be selected for a set of observations, the resulting forecast, the mean and variance of the lead time demand, and the distribution that is used to model the lead time demand.

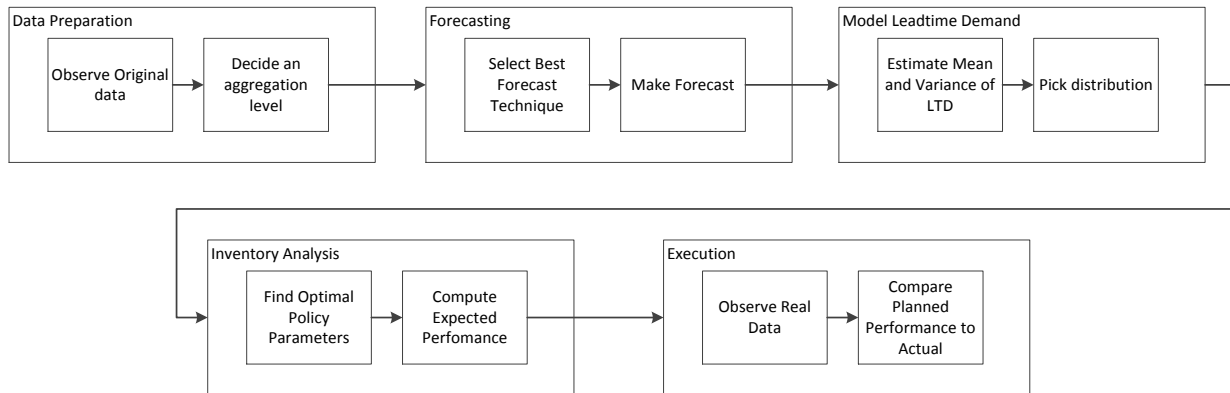


Figure 1: Aggregation Chain of Dependence

### 3.5 Measuring Effects of Temporal Aggregation

Since the level of aggregation can affect the forecast and all subsequent inventory modeling processes, the key issue that we are interested in studying is the difference between planned and actual performance. It is important to carefully consider what needs to be observed in order to measure this difference.

Let us consider this in terms of the ready rate  $\overline{RR}$ . As we can see, the computation of the ready rate in equation 5 is functionally dependent upon  $\tau$ , the aggregation level, since the parameters of  $F_{LTD}$  are dependent on  $\tau$ . In addition, the optimal values of  $r$  and  $Q$  are also dependent on  $\tau$ . Let us represent this dependence as  $\overline{RR}(\tau)$ . For example  $\overline{RR}(\tau = 1)$ , and  $\overline{RR}(\tau = 7)$  represent the predicted ready rate under daily and weekly aggregation. It is important to note that they may not necessarily have the same value. Consider the total cost in the same manner,  $\overline{TC}(\tau)$ , which represents the total cost under aggregation level  $\tau$ . A naive method for evaluating the effect of  $\tau$  would be to compute  $\overline{TC}(\tau_1)$  and  $\overline{TC}(\tau_2)$ , where  $\tau_1$  and  $\tau_2$  represent two different aggregation levels. One might be tempted to choose the aggregation level that results in the smallest total cost. However, the smallest value of  $TC$  may not be achievable during execution. The aggregation level that should be recommended (used for the item) should be the aggregation level that results in a predicted total cost that is closest to the reality of execution. Let  $TC^e$  be the total cost when the model is executed. Thus, the difference between  $TC^e$  and the predicted  $TC$  under different aggregation levels is of interest:

$$\theta_e^{TC} = TC^e - \overline{TC}(\tau)$$

We may also be interested in the relative difference:

$$\theta_r^{TC} = \frac{TC^e - \overline{TC}(\tau)}{TC^e}$$

These metrics can also be defined in a similarly for all the other major inventory performance measures,  $\theta_e^{RR}, \theta_r^{RR}, \theta_e^I, \theta_r^I, \theta_e^B, \theta_r^B, \theta_e^{OF}, \theta_r^{RR}$ . We should prefer the aggregation level that results in minimal values for these differences. In the experimental analysis, we cannot easily observe the actual execution of the system, thus we use simulation to estimate  $TC^e, RR^e, OF^e, I^e, B^e$  and form estimates for discrepancy metrics. The next section discusses the simulation model, experimental setup and associated results..

## 4 SIMULATION MODEL, EXPERIMENTAL SETUP, AND RESULTS

### 4.1 Simulation Model.

For this paper, we developed an  $(r, Q)$  inventory simulation model based on the Java simulation Library (JSL). The general capabilities of the JSL are described in Rossetti (2008). The supply chain and inventory modeling capabilities of the JSL are described in Rossetti et al. (2008). The key components of the simulation model are the demand generator and the  $(r, Q)$  inventory system model. This section provides and overview of the model and how it was constructed.

Galmes and Puigjaner (2003) introduce a flexible discrete-time arrival process, the so-called batch-on/off model, which is an extension of the single on/off source used in the case of ATM packet generation modeling. In their model, a group or batch of arrivals may be generated at any active period. Their model is characterized by a busy state, an idle state and batch size. The demand generator used in this paper, generalizes the model of Galmes and Puigjaner (2003) to a continuous time model. The generalization results in an alternating process, where the state of the process can either be in a period of low demand or a period of high demand. The period of low demand is similar to the off (idle) period in the batch on/off model, except this model allows for demand generated during the low period to be something other than zero. The period of high demand is similar to the on (busy) period. A time between demand distribution governs the rate of occurrence of demand events. This can be an arbitrary distribution. If a customer arrival occurs, the amount of the demand will be generated from  $D^H \sim G_H$  for periods of high demand and from  $D^L \sim G_L$  for periods of low demand. We assume that the low period demand is zero for simplicity. The alternating low and high periods act as filters that select the amount of demand to be generated at an arrival event. Thus, arbitrarily complex arrival processes can be generated using this methodology.

In order to set up the demand generator model, one must specify the idle state distribution ( $F_I$ ), the busy state distribution ( $F_B$ ) and non-zero demand distribution  $G(Y_{NZ})$ . In addition, the parameters of these distributions must also be provided. Any general distribution can be used to generate the values of idle period length ( $I_j$ ), busy period length ( $B_j$ ) and non-zero demand size ( $Y_{NZ}$ ). The continuous time batch on/off model is set up by specifying the mean and variance values of idle period, busy period lengths and non-zero demand size (i.e.  $\mu_I, \sigma_I, \mu_B, \sigma_B, \mu_{NZ}, \sigma_{NZ}$ ). For simplicity, each random variable is modeled with the gamma distribution to generate the corresponding values.

The JSL was used to construct an  $(r, Q)$  inventory simulation. The generated demands are processed in a manner that is consistent with the underlying theory of the  $(r, Q)$  inventory model. Since the given demand generator results in discrete demand, it is possible to apply demand-splitting. By doing so, the received (and backordered) demands can be partially or fully satisfied from the available stock on-hand. When a demand occurs, the amount demanded is determined, and then the system checks the availability of stock. If the stock on-hand is sufficient for the order, the demand is filled and the quantity on-hand is decremented. On the other hand, if the stock on-hand is not sufficient to fill the order, part of the order is backordered and the rest is filled. Therefore, based on these assumptions, a customer with a demand for multiple units is conceptually the same as multiple demands occurring at the same event time. Thus, there

is no undershoot of the reorder point. This allows the simulation model to face the compound demand process as if it faces a unit demand process (that occurs at the same event time), which allows comparable results of inventory performance measures with analytical formulas. The backorders are accumulated in a queue and they are filled on a FIFO basis after the arrival of the replenishment order. The inventory position is checked each time after a regular customer demand and the occurrence of a backorder. When the inventory position falls under the reorder point, a replenishment order is placed. The replenishment order will take an amount of time to arrive, backorders are filled, and the on-hand inventory is incremented.

## 4.2 Experimental Set-up

The system simulated is based on the attributes of a healthcare supply chain for a large healthcare integrated delivery network (IDN). The model parameters are based on the characteristics associated with a set of stock-keeping-units (SKUs) from the operating rooms in this network. The holding costs and ordering costs for each of the facilities in the network are based on an activity based cost analysis that was performed for each of the hospitals and distribution centers in the network. Like the IDN, the simulation model assumes a lead time delay of 1 day for all SKUs. The actual demand data from the set of SKUs in the IDN was utilized as inputs for the demand generator for the simulation. The following table summarizes the descriptive statistics associated with the input parameters to the simulation model. Looking at Table 1, one sees that the demand is highly intermittent as evidenced by the high variability in demand size variance (mean 5.99 and StDev 17.65) as well as the high variance associated with the idle period (mean 430.2 and StDev 2030.6).

Table 1: Summary Statistics of Input Data

Variable	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Non Zero Demand Mean	2.75	2.37	1.00	1.40	1.96	3.14	27.18
Non Zero Demand Variance	5.99	17.65	0.00	0.46	1.41	5.39	279.38
Idle Period Mean	10.97	19.93	1.01	1.86	4.19	10.05	157.00
Idle Period Var	430.20	2030.60	0.00	2.10	17.60	107.70	29316.30
Busy Period Mean	2.00	8.00	1.01	1.07	1.20	1.69	149.33
Busy Period Var	72.80	1270.80	0.00	0.10	0.30	2.00	24314.30
Unit Cost	65.31	125.43	0.95	3.04	6.79	56.09	684.40

## 4.3 Results and Conclusions

The first general observation we notice when analyzing the simulation results is that as expected, when aggregating, the demand pattern for the SKU population becomes smoother and less lumpy and intermittent. Demand classification versus aggregation level is shown in Figure 2, where (E = erratic, I = intermittent, L = lumpy, and S = smooth). Figure 2 shows how the demand pattern changes depending on the level of aggregation. When demand is aggregated on a daily basis it is intermittent and lumpy, aggregating on a weekly basis we see a significant percentage of the SKUs take on a smooth demand appearance with the a decreasing percentage of SKUs being characterized as lumpy. Finally, at a monthly level of aggregation, the majority of SKUs follow a smooth demand process. This is important because the demand characteristics have an impact on the type of forecasting model that is used for each of the SKUs.



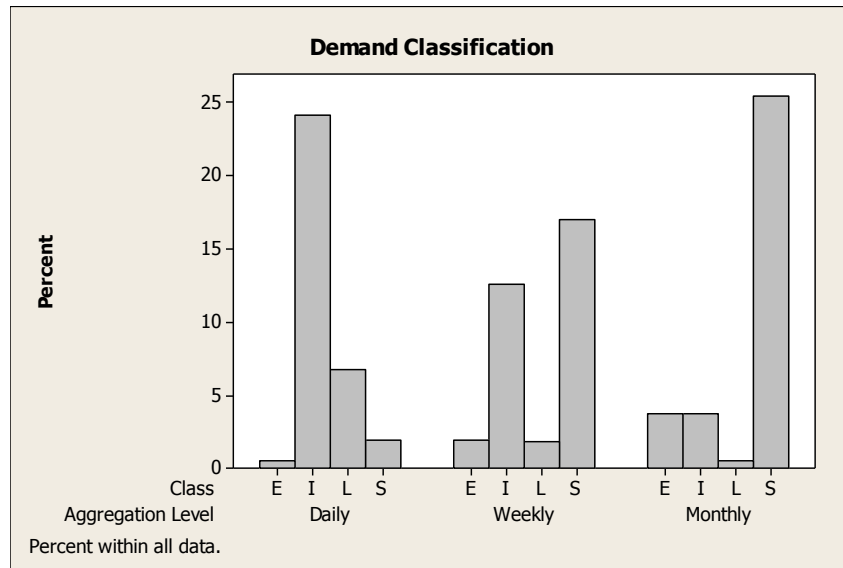


Figure 2: Demand Characteristics vs. Aggregation Level

Table 2: Summary of Relative Error for Supply Chain Performance Measures (TC = total cost, FR = fill rate, B = expected backorders, I = inventory on-hand, D = daily, W = weekly, M = monthly)

Statistic	TC			FR			B			I		
	D	W	M	D	W	M	D	W	M	D	W	M
Min	-0.69	-27.16	-0.44	-0.16	-0.41	-0.40	-2.70	-22.84	-5556.	-0.02	-0.10	-0.32
Mean	0.02	-0.12	-0.05	-0.004	0.001	-0.001	0.469	-0.35	-16.88	0.082	0.095	0.088
Max	0.93	0.85	0.89	0.008	1.000	0.015	1.000	1.000	0.988	0.504	1.000	0.504
Q1	-0.11	-0.20	-0.17	-0.00	-0.002	-0.001	0.282	-0.535	-1.785	0.026	0.030	0.027
Q3	0.117	0.015	0.006	0.000	0.003	0.004	0.794	0.381	0.232	0.105	0.114	0.109
Std. Dev.	0.200	1.434	0.204	0.011	0.057	0.023	0.453	1.784	290.37	0.081	0.105	0.098

Next, we explore the impact of using an analytically derived  $(r, Q)$  policy for each of the aggregation levels on several supply chain performance parameters. Error is defined as the difference between the simulation results and the analytical results. The simulation results estimate what will happen in the future if the  $(r, Q)$  parameters recommended by the model are used. The  $(r, Q)$  parameters were used to ensure that a minimum lower bound on the fill rate is achieved. This is a critical parameter in the healthcare setting, not having a needed SKU in the operating room can be a matter of life or death depending on the item. On average, see Table 2, the results indicate that error for total cost (TC) and fill rate (FR) are negative, which implies that the predicted analytical performance is higher than what is realized in the simulation model. This means that the analytical models are conservative with regard to fill rate which is important in the healthcare setting. Figures 3 and 4 summarize the relative error versus aggregation level for total cost and fill rate.

In Figure 3, relative error for total cost vs. aggregation level, the analytical planning model predicts a higher cost than was actually achieved for the weekly and monthly levels of aggregation. For the daily level of aggregation, the average relative error is nearly zero error. However, the interquartile relative errors for each aggregation level vary. The daily level has a 0.2318 interquartile, the weekly level has a 0.2134, and the monthly has 0.173 interquartile. This implies that the monthly aggregation level produces more consistent results and reliable actual performance levels. One last item which might distinguish weekly from monthly is the range of error in both cases, which is 28.026 for weekly and 1.367 for month-

ly. In summary, the negative cost relative error can be interpreted as a good result when using the analytical planning model. There is less risk due to cost based on using the analytical planning model. However, since the analytical planning model predicted higher cost, an organization would be “allocating” more money than needed from a planning perspective. Future work should capture the potential economic loss associated with not using those resources in other areas of the organization. In summary, these results suggest that aggregating at a monthly level and using the analytical planning model allows us to determine effective supply chain policies..

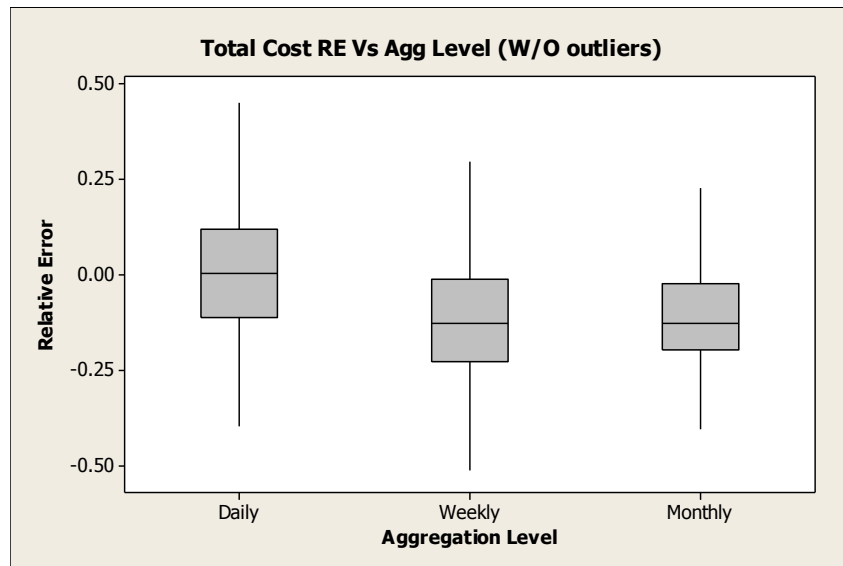


Figure 3. Total Cost Relative Error vs. Aggregation Level

From a fill rate perspective, Figure 4 illustrates that the daily aggregation level provides the best possible relative error. In this case, unlike the previously discussed total cost situation, the negative error is not welcome since it means the actual performance will be worse than expected by having a lower fill rate than predicted and required. As mentioned earlier, this is a critical concern in the healthcare supply chain setting. Something of interest is that the inter-quartiles for the three aggregation levels are very similar. The daily interquartile is 0.005, the weekly is 0.007, and the monthly is 0.007 which are all lower than 1%. This implies that the variability in the fill rate is significantly smaller than the total cost. Thus, the analytical model is unable to accurately predict fill rate and customer service level when demand is aggregated. Other approaches need to be investigated to accurately capture the fill rate should aggregation be used for planning. In summary, aggregation causes more negative relative error and thus increases a false sense of security about the level of customer service.

## 5 SUMMARY AND FUTURE RESEARCH

In this paper we explore the impact that the demand aggregation level has on our ability to forecast and establish appropriate inventory ( $r, Q$ ) policies. As expected, our results indicated that aggregation makes the demand process more smooth and makes it easier to use simpler forecasting models to predict demand. We saw that using aggregation in conjunction with the analytical model does a reasonable job of predicting total cost resulting in negative relative errors. For fill rate, we discovered that using the analytical model in conjunction with aggregation results in significant negative relative error. This presents a potential problem because the policies derived from the aggregate data give a false sense of security in terms of meeting the desired fill rate. Additional work is necessary to study the trade-offs between aggregation and policy selection and its impact on fill rate within a more robust experimental design.

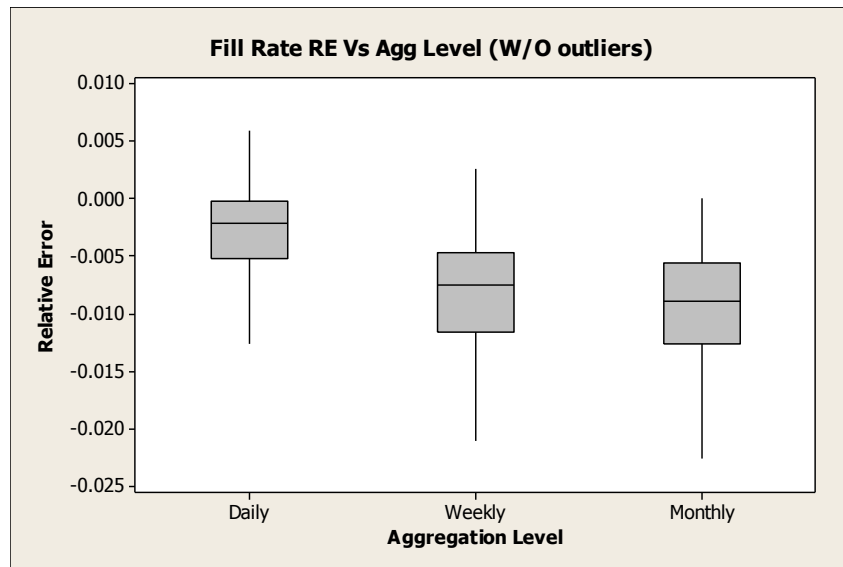


Figure 4: Fill Rate Relative Error vs. Aggregation Level

## REFERENCES

- Axsater, S. 2006. *Inventory Control*, 2<sup>nd</sup> edition, Kluwer-Academic Press
- Babai, M. Z., M. M. Ali, and K. Nikolopoulos. 2012. "Impact of Temporal Aggregation on Stock Control Performance of Intermittent Demand Estimators: Empirical analysis." *Omega*, 40: 713-721.
- Biller, B., and B. L. Nelson. 2005. "Fitting Time-Series Input Processes for Simulation." *Operations Research*, Vol. 53, No. 3, pp: 549-559.
- Cario, M. C., and B. L. Nelson. 1996. "Autoregressive to anything: Time-series input processes for simulation." *Operations Research Letters*, 19: 51-58.
- Eaves, A.H.C., and B.G. Kingsman. 2004. "Forecasting for the Ordering and Stock-Holding of Spare Parts." *Journal of the Operational Research Society*, 55:431-437.
- Galmes, S., and R. Puigjaner. 2003. "Correlation Analysis of Discrete-Time Flexible Arrival Processes." *Computer Networks*, 41: 795-814.
- Gardner, JR., E. S. 1990. "Evaluating Forecast Performance in an Inventory Control System." *Management Science*, 36, 4.
- Krajewski, L. J., L. P. Ritzman, & M. K. Malhorta. 2010. *Operations Management*. 9th ed. Prentice Hall
- Nahmias, S. 2009. *Production Operations Analysis*, 6<sup>th</sup> Edition, McGraw-Hill.
- Nikolopoulos, K., A. A. Syntetos, J. E. Boylan, F. Petropoulos, and V. Assimakopoulos. 2010. "An Aggregate-Disaggregate intermittent Demand Approach (ADIDA) to Forecasting: an Empirical Proposition and Analysis." *Journal of the Operational Research Society*, 1-11.
- Rossetti, M.D. 2008. "Java Simulation Library (JSL): An Open-source Object-Oriented Library for Discrete-event Simulation in Java". *International Journal of Simulation and Process Modelling*, vol. 4, no. 1, pp. 69-87.
- Rossetti, M. D., Varghese, V., Miman, M., and Pohl E. (2008) "Simulating Inventory Systems With Forecast Based Policy Updating", *The Proceedings of the 2008 Winter Simulation Conference*, S. J. Mason, R. Hill, L. Moench, and O. Rose, eds., Piscataway, New Jersey: IEEE.
- Sani, B., and B. G. Kingsman. 1997. "Selecting the Best Periodic Inventory Control and Demand Forecasting Methods for low Demand Items." *Journal of the Operational Research Society*, 48: 700-713.
- Silver, E. A., Pyke, D. F., and Peterson, R. 1998. *Inventory Management and Production Planning and Scheduling*, 3rd Edition, John Wiley & Sons, New York, New York

- Silver, E. A., and m. R. Rahnema. 1986. "The Cost Effects of Statistical Sampling in Selecting the Reorder Point in a Common Inventory Model." *J. of the Operational Research Society*, 37, 7: 705-713.
- Silver, E. A., and m. R. Rahnema. 1987. "Biased Selection of the Inventory Reorder Point when Demand Parameters are statistically Estimated." *Engineering Costs and Production Economics*, 12: 283-292.
- Strijbosch, L. W. G., J. J. A. Moors, and A. G. de Kok. 1997. "On the Interaction Between Forecasting and Inventory control." Technical Report No. 742, Department of Economics, Tilburg University, Tilburg, The Netherlands.
- Syntetos, A. A., and J. E. Boylan. 2006. "On the Stock Control Performance of Intermittent Demand Estimators." *International Journal of Production Economics*, 103: 36-47.
- Zipkin, P.I H. 2000. *Foundations of Inventory Management*. Boston: McGraw-Hill.

## AUTHOR BIOGRAPHIES

**MANUEL D. ROSSETTI** is a Professor in the Industrial Engineering Department at the University of Arkansas. He received his Ph.D. in Industrial and Systems Engineering from The Ohio State University. His research and teaching interests are in the areas of simulation modeling, logistics optimization, and inventory analysis applied to manufacturing, distribution, and health-care systems. He serves as an Associate Editor for the *International Journal of Modeling and Simulation* and is active in IIE, INFORMS, and ASEE. He served as co-editor for the WSC 2004 and 2009 conference, is the Publicity Chair for the WSC 2013 Conference, and will be the 2015 WSC Program Chair. He can be contacted at [rossetti@uark.edu](mailto:rossetti@uark.edu) and <http://www.uark.edu/~rossetti/>.

**MOHAMMAD A. SHBOOL** is a PhD student in the Industrial Engineering Department at University of Arkansas. He received his master's degree in industrial engineering Jordan University of Science & Technology, Jordan. He is interested in doing research in the area of Healthcare systems engineering using industrial engineering concepts, mainly inventory analysis, simulation modeling, operations research and optimization. He worked as a planning engineer at SWAROVSKI group before pursuing his PhD. He can be contacted at [mshbool@uark.edu](mailto:mshbool@uark.edu) or [m.shbool@gmail.com](mailto:m.shbool@gmail.com).

**EDWARD A. POHL** is a Professor in the Department of Industrial Engineering at the University of Arkansas. Ed retired from United States Air Force in the grade of Lieutenant Colonel and served in a variety of engineering, analysis and academic positions during his career. Ed received his Ph.D. in systems and industrial engineering from the University of Arizona, an M.S. in reliability engineering from the University of Arizona, an M.S. in systems engineering from AFIT, an M.S. in engineering management from the University of Dayton, and a B.S.E.E. from Boston University. His primary research interests are in repairable systems modeling, supply chain modeling, risk and reliability, decision making, engineering optimization, and probabilistic design.

**VIJITH M. VARGHESE**, Ph.D. is a Senior Engagement Manager in the BankSmarts Solutions and has around ten year experience in design and analysis of supply chain and logistics problems from banking, healthcare, retail, and military systems using Operations Research and Statistical Modeling. He has been a Postdoctoral Fellow for the Center for Innovation in Healthcare Logistics. He has been a researcher for numerous projects for US Air Force Office of Sponsored Research, Center for Excellence in Logistics and Distribution, US Naval Supply Systems Command, VHA Inc., Covidien Inc., Mercy Health System, Mayo Clinic, and Wal-Mart. He holds a Ph.D. in Industrial Engineering and M.S. in Industrial Engineering from the University of Arkansas and a bachelor's degree in Mechanical Engineering. He is an active member of Alpha Pi Mu, IIE and INFORMS and has more than 25 publications in leading journals and proceedings.