ELSEVIER

Int. J. Production Economics 0 (■■■■) ■■■–■■■

international journal of production economics

www.elsevier.com/locate/dsw

A robustness study of a multi-echelon inventory model via simulation

Yeu-San Tee*, Manuel D. Rossetti

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

Received 1 January 2002; accepted 20 May 2002

Abstract

This paper examines the robustness of a standard model of multi-echelon inventory systems, specifically the models discussed in Axsater (Oper. Res. 48(5) (2000) 686). A simulation model was developed to explore the model's ability to predict system performance for a two-echelon one-warehouse, multiple retailer system using (R, Q) inventory policies under conditions that violate the model's fundamental modeling assumptions. In particular, the impact of non-stationary demand on this stationary demand inventory model was examined. The model performs well at the low demand and large retailer order batch size situations, but our testing of the model indicated that care must be taken when applying this model to situations that violate its fundamental assumption. These results should help practitioners to better understand the assumptions of these models and to determine when or when not to apply these models in practice. © 2002 Published by Elsevier Science B.V.

Keywords: Robustness; Non-stationary demand; Multi-echelon inventory

1. Introduction

Efficient and effective management of inventory throughout the supply chain can significantly improve customer service levels and reduce system cost. During the last decade, previous research has led to the development of many analytical inventory models, which can be embedded in decision support systems to assist in inventory management, such as those that are used in the Enterprise Resource Planning, Supply Chain Management, and Advanced Planning System software tools. Most of the time, these models and systems are treated as a black box in obtaining solutions. Uncertainty in the real world could cause the misuse of the models. Practitioners should be careful when using these models and should not ignore the uncertainty in the system that could affect the model performance and cause serious consequences to a company's inventory management strategy.

These inventory models are developed and designed for a specific system based on certain conditions and assumptions. The models should be robust enough to be used under some unforeseen situations. For example, a model may assume Poisson demand for customer orders, but the actual demand pattern shows non-Poisson characteristics or seasonal trends. A robust model

*Corresponding author. Tel.: +479-575-6756; fax: +479-575-8431.

E-mail address: rossetti@uark.edu (Y.-S. Tee).

0925-5273/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S 0 9 2 5 - 5 2 7 3 (0 2) 0 0 2 5 9 - 1

should still be able to provide accurate performance prediction/approximation for the inventory system even when the actual environmental conditions have violated the modeling assumptions. Demand and lead-time are two main conditions that are easily affected by the randomness and changes to environmental conditions. Therefore, it is important to explore the robustness of inventory models, and to study the impact of violations to the demand and lead-time assumptions on the model's output. In addition, the knowledge gained on the quality of the model's outputs under various violated conditions will help determine when and when not to use these models in practice.

This research focuses on testing the robustness of a recent model of multi-echelon inventory systems via computer simulation. The study determines how the model performs under violated assumptions and the conditions where the models will perform the worst in predicting the system performance measures. The model tested considers a distribution network consisting of one warehouse and N retailers, where the retailers directly serve the customers and the warehouse replenishes all the retailers. At each warehouse and retailer location, when the inventory position (net inventory on hand plus stock on order minus backorders) drops below the reorder point R, a replenishment order batch size of Q is placed. This type of inventory policy is relatively easy to implement with the point-of-sales terminal and transaction reporting systems. Many have suggested using the continuous review (R, Q) inventory control policy on the slow moving type A items (Silver et al., (1998), Hopp and Spearman (2000), Zipkin (2000), Axsater (2000), etc.).

Many models have assumed a probability distribution with known parameters to represent the demand process. The stationary Poisson distribution has been widely used to model the demand in inventory models; however, seasonal type items, short product life cycles, and volatility in the marketplace suggest that the probability distribution of demand tends to change over time, i.e. the demand is non-stationary. This paper will examine the effects of violating the stationary Poisson demand assumption of the model presented in Axsater (2000) with a simple non-stationary Poisson demand process.

The next section provides an extensive review of relevant literature, and is followed by the simulation methodology and the experimental design used in during our analysis. Then, the summary of the experimental results with discussion is presented. The last section concludes this paper with recommendations and directions for future research.

2. Literature review

Many early multi-echelon inventory models have been used for military contingency support. For example the work of Sherbrooke (1968) and Muckstadt (1973) discusses how to control repairable items in military base-depot supply systems. Since then, many other multi-echelon inventory systems have been studied extensively, especially for the service part inventory control system and for the one-for-one base stock ordering policy (a special case of the (R, Q) policy with $Q = 1$). Of note is the work of Cohen et al. (1986) which has been successfully integrated into IBM's OPTIMIZER, a multi-echelon service inventory optimization software support system, as described by Cohen et al. (1990). IBM reported a savings of over \$250 million resulting from the use of their OPTIMIZER software.

The continuous review (R, Q) policies two-echelon system has also received tremendous attention. A review of the development of the two-echelon (R, Q) continuous review inventory models is presented to motivate the identification of a significant model for testing in this study. Lastly, previous research related to the robustness study of inventory models is presented.

2.1. Two-echelon (R, Q) inventory models

A good review of the models dealing with continuous review policies for multi-echelon inventory systems can be found in Axsater (1993a). The traditional method focuses on the steady-state behavior of the inventory levels, where the lead-time demand is approximated by the mean and variance incorporated into a certain distributional form. The multi-level system is decomposed into single locations to be evaluated separately with parameters that depend on each other. The total cost function is obtained through the average inventory and backorder units. Pioneering research in this approach is Deuermeyer and Schwarz (1981), Moinzadeh and Lee (1986), Lee and Moinzadeh (1987), and Svoronos and Zipkin (1988). Svoronos and Zipkin (1988) provided several refinements and extensions to the work developed in Deuermeyer and Schwarz (1988). The approximation method of Svoronos and Zipkin (1988) has been shown by Axsater (1993a, b) to be accurate for the identical retailer case.

As opposed to the traditional approach, Axsater (1993a, b) suggests several approximation methods. These approximations are derived based on the previous work of Axsater (1990), which used a recursive procedure for the evaluation of one-warehouse and N retailers with one-for-one policies. The holding and backordering costs must be functions of the delay experienced by the customer. His numerical results show that his approximations provide good results that are comparable to that of Svoronos and Zipkin (1988). Following this line of research, Axsater (1998), Forsberg (1996), and Forsberg (1997) developed models to evaluate the non-identical retailers case.

Since previous models by Axsater and Forsberg are based on the weighted average costs for one-for-one policies, the models are limited to pure Poisson demand processes only and rely on a special cost structure. In Axsater (1995), he began to investigate the steady-state behavior of the inventory levels, and in Axsater (2000), he provides an exact analysis through determining the complete probability distributions of the retailer inventory levels in steady state. This model uses a common cost structure and the model can be used to solve the one-warehouse and non-identical retailer case with compound Poisson demand. This model is considered the state-of-art in exactly evaluating two-echelon inventory systems with continuous review (R, Q) batch ordering policies for the low demand items such as spare parts. Therefore, the Axsater (2000) was selected for testing in this study.

2.2. The robustness of inventory models

Some previous studies have been performed to test the robustness of inventory models directly and indirectly. Many studies are performed on single location inventory models. The work of Naddor (1978), Fortuin (1980), Banks and Spoerer (1986), and Lau and Zaki (1982) concludes that the optimal inventory decisions of single location inventory models are affected more by the means and standard deviations of the demands rather than the form of the demand distribution. Tyworth and O'Neill (1997) examine the use of the normal approximation in determining the safety stock for the (R, Q) continuous review inventory models. By comparing the solutions (safety stock, total cost and fill rate) from the normal approximations and exact approaches, they found that the normal approximation method is robust across seven industry groups (fast-moving demand items). Nevertheless, Fotopoulos et al. (1988) present a method to determine the safety stock when the demands are autocorrelated and the lead-times are random. Their numerical results show that ignorance of autocorrelation in demand could provide severe errors when determining the safety stock; however, the effect of non-normal demand was found to be relatively small.

The above studies were all concerned with single-echelon inventory models. The only study that we found which examined the behavior of multi-echelon models under violated model demand assumptions was performed by Lagodimos et al. (1995). They tested the robustness of two two-echelon (serial) periodic review order-up-to-S inventory models. The independent identically distributed normal demand assumptions were violated with stationary autocorrelated demand processes. Lagodimos et al. (1995) performed their analysis using an analytical approach in which the model parameters were redefined to fit any stationary autocorrelated normal demand process in an exact closed form. Their results demonstrate that the models, when ignoring the effects of autocorrelation, might provide significant error in predicting system performance depending on the overall system parameters settings.

Tee and Rossetti (2001) proposed using simulation to evaluate the robustness of a two-echelon inventory model. The model presented in Axsater (2000) was tested under stationary non-Poisson demand and non-constant lead-time conditions using simulation methods. The effects of the variability of the demand and the lead-time were evaluated via Gamma distributed time between demands and lead-time with different coefficients of variation. The results showed that if the testing conditions are ignored when using the analytical multi-echelon inventory model, then significant error in predicting the inventory system performance can be obtained. The testing conditions influenced the prediction of the number of backorders the most, and the number of backorders tends to be under-predicted due to the increase of demand and lead-time variability. The error in predicting the expected number of backorders is the main factor which influences the errors in predicting the total system cost, and the error is as much as 34% over and 40% under the actual system cost.

Considering an item that uses exponentially weighted moving average demand forecasting techniques, Graves (1999) incorporated non-stationary demand into a single-item inventory model, and further extended the model into a two-stage inventory system. Based on numerical observations, more safety stock was needed when the demand is non-stationary. He also observed that the demand process at the upper stage is more variable when the downstream stage experiences non-stationary demand, which is the so-called *bull-whip effect*.

Based on this review of the literature, we conclude that while much research has been done on multi-echelon inventory models, the robustness of these models remains to be examined. Many models assume simple stationary demand processes. The typical time dependent demand found in practice should not be forgotten when using these inventory models. Thus, we examine if significant errors in estimating the inventory system performance will occur under the non-stationary demand circumstances and the overall robustness of these models.

3. Methodology

The purpose of this study was to assess the quality of the model presented in Axsater (2000) based on how the model's prediction of inventory system performance compared to the true value when the assumptions are violated. A simulation model was used to provide the true system performance under the violated modeling assumptions. Our research methodology is as follows: First, the Axsater (2000) model is used to obtain the recommended optimal policies and the predicted values of the system performance measures. Then, a simulation model was developed to incorporate the model testing conditions. The simulated performance measures are compared to the model's predicted values under a design of experiments for robustness analysis.

3.1. Problem setting

The Axsater (2000) model was obtained in a prototype program available via contact with Sven Axsater (E-mail: sven.axsater@iml.lth.se). We decided to analyze only the identical retailer case for a simpler and clearer analysis of the two-echelon system. Previously, Svoronos and Zipkin (1988) provided an accurate approximation for the identical retailer system with 32 test problems. In these 32 test problems, the demand was assumed to be a stationary Poisson process at the retailer level, and all the holding cost factors ($h_r = h_w$) are \$1 per unit and transportation lead-times are 1 day for all cases. The factors and levels of these 32 test problems are shown in the Table 1.

We used the Axsater (2000) program to solve the 32 test problems, and the inventory system performance estimation and optimal inventory policies obtained are not significantly different from the results given in Svoronos and Zipkin (1988). The performance measures of interest are the expected inventory on-hand across all the retailers (I_r), the expected number of backorders across all the retailers (B_r), the expected inventory on-hand at the warehouse (I_w), and the expected total system cost. The total cost equals warehouse inventory holding cost (WIC) + all retailers inventory holding cost (RIC) + all retailers

Table 1
Experimental factors and levels for the 32 test problems

Factors	Symbols	Levels	Level
Average demand	D	2	Low—0.1 unit demand per period High—1.0 unit demand per period
Number of retailers	N	2	Small—4 retailers Large—32 retailer
Backorder cost factor	P_r	2	Small—$5 per unit backordered Large—$20 per unit backordered
Retailer order quantity	Q_r	2	1 unit 4 units
Warehouse order quantity	Q_w	2	1 Q_r batch-unit 4 Q_r batch-units

backordering cost (RBC), where the WIC, RIC, RBC are obtained from the I_w, I_r, and B_r with the holding cost factors ($h_r = h_w$), backordering cost factor (P_r), and the number of retailers (N). All these values are used as the baseline for testing and comparison. The recommended inventory policies (the reorder points of warehouse and retailers) are used as simulation inputs.

3.2. Simulating the non-stationary Poisson demand

A simulation model was developed to represent the two-echelon inventory system represented by the analytical model. Tee and Rossetti (2000) presents the details of the simulation model including the logic, structure, data inputs, outputs, verification, and validation. The simulation model built in this study is simple and easy to use, and the model can be modified to accommodate other distribution systems. A single location model was first built and then expanded into a warehouse-retailer model in Arena 5.0 Professional Edition. The simulation models were verified and validated to give performance measures that are an accurate and valid representation of the system.

The main testing condition for this study is the non-stationary Poisson demand process. A piece-wise-constant arrival function with two rates over the year was chosen to model the non-stationary characteristics of the demand pattern. The demand process is assumed to have a lower than average demand for the first half-year, and higher than average demand for the second half-year, and then it repeats in a yearly cyclical pattern. We felt that such a cyclical demand pattern would be sufficient to evaluate the performance of the stationary analytical model under the time-dependent demand situation. Furthermore, the Arena simulation software provides an easy way to generate such non-stationary Poisson demand process through the built-in 'SCHEDULES arrival' element. The method behind this SCHEDULES arrival element is via the inversion of a stationary rate-one Poisson process against the cumulative rate functions as described by Law and Kelton (2000).

3.3. Experimental design

With the simulation model and the recommended inventory policies from Axsater (2000), we examined the 32 test problems under the non-stationary Poisson demand scenarios. Table 2 shows the violated demand conditions for the two different demand rates in the 32 test problems. For example, if the test problem has an average demand rate of 1 item per day, the simulation model will run at average demand rate of 0.5 items per day for the first 6 months, and then 1.5 items per day for the next 6 months. This represents a time-weighted yearly demand of 1 item per day.

According to Needham and Evers (1998), an inventory system is a non-terminating system and one must design the experiment to evaluate the

Table 2
Non-stationary Poisson demand process

Piece-wise constant rates:				
2 periods per year	Demand ($D = 0.1$)		Demand ($D = 1.0$)	
	Rate per day	Average	Rate per day	Average
Period #1	0.05	0.1	0.5	1
Period #2	0.15	0.1	1.5	1

system under steady-state conditions. The experimental design analysis of a simulation model must provide sufficient independent observations to do statistical tests and to obtain statistical significance. The batch means method was used for this steady-state estimation. Initial inventory on-hand, on-order, and backordered at each location were set to zero for the simulation model, and these conditions could cause initialization bias. Welch's plot procedure as described in Law and Kelton (2000) was used to determine the warm-up period.

Since the performance measures for the inventory system will not be independent of time when the demand process is non-stationary, statistical analysis for steady-state cycle parameters estimation was used. A discussion of steady-state cycle parameter estimation is available in Law and Kelton (2000). Since the non-stationary pattern repeats on a yearly cycle, the cycle was determined to be 1 year and observations were collected for each year. For example, we let Y_i be the average inventory level during year i. We are interested in estimating the steady-state mean of Y_i, $E[Y_i]$. The steady-state distribution of the *yearly* performance should still exist even though the system is non-stationary. No initialization bias was observed when estimating the steady-state mean performance on a yearly basis so that no warm-up period was needed for this non-stationary demand simulation. Thirty batches of 1 year was found to be enough to ensure independent and identically distributed data for a 95% confidence level and a confidence interval width of 1.41 for the average annual total cost measurement was obtained. This design of experiments retains the original 2^5 factorial experiment (the 5 factors in Table 1) with 32 design points.

To measure the robustness of the model in predicting the performance of the system, the deviation (error) of the model-predicted values from the actual true values (simulated values) are needed for comparisons. If the error is a positive number, the model over-predicts the system, which means the actual value is lower. If the error is negative, the system performance is under-predicted implying that the actual value is higher. Since each test case has different conditions with certain recommended inventory policies, the performance measure values are not the same for each case. The error will show how much the prediction is off from the actual value, but it will not tell how sensitive the differences are. The relative error is a response that can be used to indicate the relative differences. The relative error is defined as the following:

$$\text{Error} = \text{Model Prediction Value} - \text{Actual Simulated Value},$$

$$\text{Relative Error} = \text{Error}/(\text{Actual Simulated Value}).$$

Nevertheless, the relative error has a disadvantage, which is the over sensitivity of this measure. When the response results in small performance measure values, little deviation will give a high relative error. The error and relative error of the average total cost are the main responses used in the statistical analysis of the factorial experiments in this study. Analyses on other performance measures such as the average number of backorders and average inventory levels were also performed to study the tradeoff between the components of the average total cost.

4. Experimental results and discussion

Each of the 32 design points has 30 observations (replicates) from the 30-batch simulation runs, yielding a total of 960 estimates of each performance measure. Table 3 reports the summary statistics of the error and relative error in predicting performance measures under the non-stationary Poisson demand condition from all the

Table 3
Overall performance measures error and relative error summary statistics

Summary statistics	Average total cost		Average retailer backorders (B_r)		Average warehouse inventory (I_w)		Average retailer inventory (I_r)	
	Error	Relative error	Error	Relative error	Error	Relative error	Error	Relative error
Mean	−10.1231	−0.1426	−0.0465	−0.2954	−1.3053	−0.1989	0.0086	0.0045
Std. dev.	16.0065	0.1385	0.0535	0.2843	2.1167	0.2626	0.0592	0.0382
Maximum	2.2430	0.2055	0.0895	1.2727	1.9960	0.6005	0.3187	0.1957
Upper quartile	−0.4090	−0.0269	−0.0025	−0.0517	0.0000	0.0000	0.0317	0.0508
Median	−2.2750	−0.1069	−0.0233	−0.3460	−0.2200	−0.1028	0.0028	0.0025
Lower quartile	−15.8610	−0.2554	−0.0832	−0.5396	−3.0487	−0.3774	−0.0114	−0.0082
Minimum	−59.5740	−0.4638	−0.2353	−0.7713	−6.0390	−0.7722	−0.3064	−0.1950
Sample size	960	960	960	960	960	960	960	960

simulation observations (*across all experiments*). The summary statistics include the sample mean, sample standard deviation, maximum, upper quartile, median, lower quartile, minimum, and sample size.

The summary statistics demonstrate that all the performance measures except the I_r are under-predicted for the non-stationary Poisson demand conditions, i.e. all the mean errors, mean relative errors, and 75% or more of the data are negative. The overall statistics also indicate that the I_r has the least mean and smallest variation of error (mean = 0.0086, std. dev. = 0.0592) and relative error (mean = 0.0045, std. dev. = 0.0382). On the other hand, B_r relative error has the most variation with mean of −0.2954, standard deviation of 0.2843, maximum of 1.2727, and minimum of −0.7713.

In order to examine the risks associated with the model, we examined the probability that outputs from the model will be greater than 10% in absolute relative error when compared to the true value. Table 4 shows the summary of absolute relative error risk from the experiments by counting the number of times the absolute relative error was greater than 0.1 and then dividing by the sample size of 960. The table also includes the probability when the average demand rate is low ($D = 0.1$) and high ($D = 1.0$) each with 480 samples to show the significance of the demand factor. As shown in Table 4, 72.29% of the experiments had relative errors greater than 10% as compared to the true value of the B_r across both demand conditions. The percentage is even higher for the high demand case, 99.79%. Again, the I_r was determined to have the least risk as shown in Table 4, 2.71% of the experiments had a relative error greater than 10% from the true mean. In general, Table 4 indicates that the model has a high risk of performing poorly for the prediction of cost, B_r, and I_w, and the risk of poor performance is worse when the demand rate is high (more than 90% of the cost and B_r experimental values have large error).

After analyzing the overall summary statistics, exploratory plots of the average relative error values of the 32 design points (Fig. 1) were used to check for patterns in each performance measure. As shown in Fig. 1, the relative error of the cost, the B_r, and the I_w seem to have similar patterns (same signs and similar trends). Nevertheless, the relative errors of I_r are smaller (all less than 10%), and the I_r's pattern tends to have the opposite sign with respect to the other performance measures especially at the last few design points. In other words, most of the performance measures are observed to be under-predicted most of the time, but the I_r is over-predicted. The exploratory plots also show that all the errors are smaller when the demand is low (design points 1–16).

To further investigate the effects of the errors in I_w, I_r, and B_r on the total system cost, the percentage of WIC, RIC, and RBC to the total system cost for all the original cases and the experiments were computed for analysis. Please refer to Table 5 for the relationship among these inventory system performance measures, and the percentage change of the cost components. As shown in Table 5, the direction of percentage change of the RBC and RIC components from original cases to experiments are always opposite of each other, i.e. as the percentage of the RBC increased (positive % difference), the RIC percentage decreased (negative % difference). The percentage change of the WIC cost component, however, does not fluctuate as much since the percentage differences are always less than or equal to 5.5%. Based on the average percentage across all the 32 original cases, RIC is the main component of the total cost with 58%, followed by RBC with 32%, and then WIC with 10%. After the introduction of the non-stationary Poisson demand into the experiments, the RIC percentage of the total cost decreased 9–49%, RBC increased 8–40%, and WIC only increased 1–11%. RIC and RBC are still the main components for the total cost for most cases. As the B_r was consistently under-predicting more (negative error and relative error), there are more backorders in the system and hence the RBC percentage of the total cost is increased. Therefore, together with the similar patterns of the B_r and total cost error and relative errors as shown in the exploratory plots (Fig. 1), this indicates that the error in predicting the B_r is the main driver for the total system cost change for these experiments.

The main effect and interaction plots of the non-stationary Poisson demand study were used to analyze the sensitivity of the factors and interactions graphically (to evaluate the behavior of the analytical model). Most of the main factors are found to significantly affect the errors and relative

Table 4
Probability of error greater than 10% of the true value from model prediction

Performance measures	Probability (absolute relative error >10%)		
	$D = 0.1$	$D = 1.0$	Overall
Cost	0.1000	0.9646	0.5323
B_r	0.4479	0.9979	0.7229
I_w	0.2417	0.8292	0.5354
I_r	0.0500	0.0042	0.0271
Sample size	480	480	960

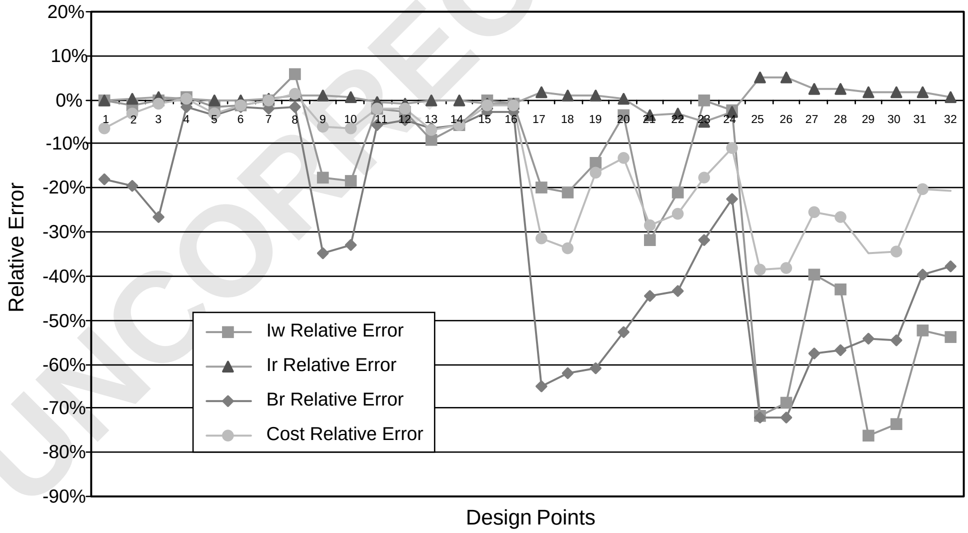

Fig. 1. Exploratory plots of the average relative errors of each response.

Table 5
Cost components percentage change from original cases to experiments

Design point	RBC/total cost			RIC/total cost			WIC/total cost		
	Original (%)	Experiment (%)	% Diff.	Original (%)	Experiment (%)	% Diff.	Original (%)	Experiment (%)	% Diff.
1	31.4	35.8	4.5	68.6	64.1	−4.5	0.0	0.0	0.0
2	13.1	15.8	2.6	64.5	62.2	−2.3	22.4	21.9	−0.4
3	4.2	5.6	1.5	95.8	94.4	−1.4	0.0	0.0	0.0
4	19.2	19.5	0.3	40.1	40.1	0.0	40.7	40.4	−0.3
5	77.8	78.1	0.3	0.0	0.0	0.0	22.2	21.9	−0.3
6	68.1	68.1	0.0	0.0	0.0	0.0	31.9	31.9	0.0
7	16.9	17.2	0.3	83.1	82.8	−0.3	0.0	0.0	0.0
8	28.9	29.8	0.9	45.6	45.9	0.2	25.4	24.3	−1.1
9	11.8	17.1	5.2	84.6	78.8	−5.8	3.5	4.0	0.5
10	14.0	19.5	5.5	83.1	77.2	−5.9	2.9	3.3	0.4
11	26.3	27.3	1.0	66.4	65.4	−1.0	7.3	7.3	0.0
12	26.1	26.8	0.7	63.9	63.1	−0.8	10.0	10.1	0.1
13	89.5	89.3	−0.2	0.0	0.0	0.0	10.5	10.8	0.3
14	87.2	87.2	0.0	0.0	0.0	0.0	12.8	12.8	0.0
15	16.9	17.1	0.3	83.1	82.8	−0.3	0.0	0.0	0.0
16	15.9	16.2	0.3	80.9	80.7	−0.3	3.2	3.1	0.0
17	23.9	46.7	22.8	64.4	43.3	−21.1	11.7	10.0	−1.7
18	29.8	52.1	22.3	60.6	39.9	−20.7	9.6	8.0	−1.5
19	12.5	26.5	14.1	79.6	65.8	−13.9	7.9	7.7	−0.2
20	13.2	24.2	11.0	68.6	59.4	−9.2	18.2	16.3	−1.9
21	42.3	54.4	12.1	48.1	35.5	−12.6	9.6	10.1	0.4
22	38.9	50.9	12.0	47.1	36.0	−11.1	14.0	13.1	−0.9
23	38.5	46.6	8.1	61.5	53.4	−8.1	0.0	0.0	0.0
24	35.5	40.9	5.4	40.2	36.9	−3.3	24.2	22.2	−2.1
25	23.0	50.8	27.8	74.3	43.4	−30.9	2.7	5.8	3.1
26	22.5	50.2	27.6	74.4	43.8	−30.6	3.0	6.0	3.0
27	23.6	41.7	18.0	70.8	51.4	−19.4	5.6	6.9	1.3
28	26.0	44.2	18.2	69.3	49.7	−19.6	4.8	6.1	1.4
29	37.2	53.0	15.9	59.7	38.3	−21.4	3.2	8.7	5.5
30	36.5	52.4	15.9	59.9	38.5	−21.3	3.7	9.0	5.4
31	32.5	42.8	10.2	62.5	49.0	−13.6	4.9	8.3	3.3
32	35.1	44.9	9.8	60.7	47.9	−12.8	4.2	7.2	3.0
Average	32	40	9	58	49	−9	10	11	1
Std. dev.	21.2	21.0	8.7	25.9	24.7	9.6	10.2	9.5	1.9
Min	4.2	5.6	−0.2	0.0	0.0	−30.9	0.0	0.0	−2.1
Max	89.5	89.3	27.8	95.8	94.4	0.2	40.7	40.4	5.5

Total cost = RBC + RIC + WIC

RBC $= N \times P_r \times B_r$ RIC $= N \times h_r \times I_r$ WIC $= h_w \times I_w$

errors from the low to high level, except the main factor Q_w. Table 6 shows the effects of factor D, Q_w, and Q_r on the relative errors. The D factor was found to be the most significant: the cost, B_r, and I_w will be underestimated more (relative error is larger) when changing from low to high demand; when the demand is high the I_r relative error is larger, but in this case I_r is overestimated more. Another important observation is the effect of the Q_r factor. Smaller relative errors are found at the larger retailer order quantity ($Q_r = 4$): cost, B_r, and I_w are underestimated less; I_r is overestimated less.

Table 6
The effects of factor D, Q_w, and Q_r on the relative errors

Factor	Relative error	Sign (±)	Main effect plots: changes of mean value		Remark
			From low level	To high level	
D	Cost	−	Under-predicted less	Under-predicted more	Smaller error at low demand
	B_r	−	Under-predicted less	Under-predicted more	
	I_r	+	Over-predicted less	Over-predicted more	
	I_w	−	Under-predicted less	Under-predicted more	
Q_r	Cost	−	Under-predicted more	Under-predicted less	Smaller error at high Q_r
	B_r	−	Under-predicted more	Under-predicted less	
	I_r	+	Over-predicted more	Over-predicted less	
	I_w	−	Under-predicted more	Under-predicted less	
Q_w	Cost	−	Under-predicted	Under-predicted	Not much change, no effect
	B_r	−	Under-predicted	Under-predicted	
	I_r	+	Over-predicted	Over-predicted	
	I_w	−	Under-predicted	Under-predicted	

To further support that the model performs well when $D = 0.1$ and $Q_r = 4$, the probability of relative error greater than 10% of the true value for all four groups of D and Q_r combinations was computed. Table 7 shows that only 5.83% of the total cost (sample size of 240) has error greater than 10% of the true value when $D = 0.1$ and $Q_r = 4$. Even though B_r has considerable high probability (33.33%) of error greater than 10% from the true mean, it is still the lowest compared to the other three categories. The I_w also has the smallest probability among the four groups of D and Q_r combination with 15.83%; however, the I_r is not being predicted that well at $D = 0.1$ and $Q_r = 4$ with a probability of 0.1 (10%) when compared with other groups.

In these experiments, similar results to the previous research in Tee and Rossetti (2001) are found. The violated assumptions affect the prediction of B_r (the expected number of backorders across all the retailers) the most, and the error in predicting B_r is the most influential factor of the total cost error. The B_r as well as the total cost tends to be under-predicted. The inventory policy recommended by an analytical model depends upon the tradeoffs between the cost components under certain assumptions. Under the conditions that have more demand variation and uncertainty than the model assumed, the model will recommend a policy of carrying less safety stock than is needed, and hence more backorders occur and the service level is reduced. If the actual demand is less variable, less backorders will happen under the recommended inventory policy, which means the model over-predicts and influences the system to carry more inventories.

The experimental results under the non-stationary Poisson demand process show that the model does not perform well when the demand process is non-stationary; however, the model is still within an acceptable range when the average demand rate is low ($D = 0.1$) and the retailer order batch size is large ($Q_r = 4$). The non-stationary Poisson demand in this study designates the year duration into two periods, one with a lower than average demand and the other with a higher than average demand, i.e. the demand variability over the year is increased. When the demand is lower than the average, the inventory system has more safety stock in the first half period, and hence the number of backorders is over-predicted. On the other hand, higher demand increases the number of backorders as the inventory system has less safety stock, which leads to under-prediction. Nonetheless, the overall results show that B_r is under-predicted in most cases, which means that the over-prediction and under-prediction during the whole year do not even out, and the effects of under-prediction are much higher.

Table 7
Probability of error greater than 10% of the true value for D and Q_r groups

Performance measures	Probability (absolute relative error $>10\%$)			
	$D=0.1$, $Q_r=4$	$D=0.1$, $Q_r=1$	$D=1.0$, $Q_r=4$	$D=1.0$, $Q_r=1$
Cost	0.0583	0.1417	0.9292	1.0000
B_r	0.3333	0.5625	0.9958	1.0000
I_r	0.1000	0.0000	0.0083	0.0000
I_w	0.1583	0.3250	0.6583	1.0000
Sample size	240	240	240	240

While the average demand is low ($D=0.1$), there is less fluctuation in the demand, and the differences between the over-prediction and under-prediction in a year are not as large. Therefore, the model does not perform poorly when considering the yearly performance (the over-prediction in the first half period and under-prediction in the other period balance each other out). The model will perform even better when the retailer order quantity is large. This is because when the order quantity is larger, the replenishment order will be placed less frequently, and hence there is less chance for backorders to occur.

The prediction of the average inventory level at the retailer is not significantly affected by the introduction of higher or lower demand variance given that the reorder point and order quantity at the retailer are pre-specified by the model. On the other hand, the error in the average inventory level at the warehouse is affected significantly by the testing conditions. A possible explanation for such a phenomenon is that the violation of the assumptions for the retailer demand processes causes the warehouse to experience less of a renewal process for replenishment orders. Therefore, the model assumptions for the warehouse will also be violated, and the error in the average inventory level becomes more significant. There is, in essence, a "bull-whip" effect that amplifies the problems caused by the violation in the assumptions to the warehouse.

5. Conclusions and future research

This study evaluated the behavior of a (R,Q) multi-echelon inventory model in predicting the total system cost under a non-stationary Poisson demand process. Assumptions such as Poisson demand may be convenient for analytical modeling, but can be inadequate for some inventory distribution systems. Neither the over-prediction nor the under-prediction is good for a company who uses these analytical models (see Table 8). If the total system cost is over-predicted, the company might hold too much capital for its distribution system; capital that could be invested elsewhere. If the total system cost is under-predicted, the company might not have enough capital to cope with the actual situation. Under-prediction is considered to cause more serious losses to the company because actual service level is deteriorated (more backorders in the system), and hence the profits may be reduced.

The simulation results of this study show that the model had significant error in predicting the total system cost (tends to be under-predicted) when the actual demand is non-stationary Poisson. The cost relative error ranged between 21% and −46%. The error in predicting the backorders was found to dominate the effects on the total system cost with the relative error ranged between 127% and −77%. The prediction of the average inventory level at the warehouse by the model is also shown to be not accurate, but its effect on the total system cost is not large because the warehouse inventory cost is only a small portion of the total system cost. Moreover, the results find that the model gives less error in predicting the average inventory level at the retailer.

As a conclusion, the (R,Q) two-echelon inventory model considered in this study has potentially serious risks involved if used under non-stationary Poisson demand conditions. Companies should be

Table 8
Consequences of under-predicting and over-predicting the system

What if?	Model prediction under violated conditions		
	Accurate	Under-predicting	Over-predicting
Use the model	Good	Under-planned; service level is deteriorated	Over-planned; unnecessary capital allocation

aware of the possibility of non-stationary or seasonality of the demand while using analytical inventory model designed for stationary situations. Although the model performs well at the low demand and large retailer order batch size situations, the robustness of the model over all conditions tested is in doubt. We recommend that practitioners evaluate the potential use of multi-echelon methodologies in a simulation study before implementation.

In this study, the effects of the violated conditions on the optimal inventory decisions were not investigated. The optimal policies recommended by the model would more likely be sub-optimal or not optimal under the uncertain conditions. We could examine ways to use the model even though the model assumptions are violated. For example, an algorithm can be designed to re-optimize the inventory policies when the demand condition changes, i.e. a more efficient way to control the inventory by an adaptive procedure. Treharne and Sox (1999) have reviewed models and methods for adaptive inventory control. There are potentially significant cost savings and benefits with the use of adaptive control policies under changing environmental condition. It is our hope that this research will spark an interest in developing more robust models for inventory control.

References

Axsater, S., 1990. Simple solution procedures for a class of two-echelon inventory problems. Operations Research 38 (1), 64–69.

Axsater, S., 1993a. Continuous review policies for multi-level inventory systems with stochastic demand. In: Graves, S.C., et al. (Eds.), Handbooks in OR and MS, Vol. 4, pp. 175–197.

Axsater, S., 1993b. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. Operations Research 41 (4), 777–785.

Axsater, S., 1995. Approximate evaluation of batch-ordering policies for a one-warehouse, N non-identical retailer system under compound Poisson demand. Naval Research Logistics 42, 807–819.

Axsater, S., 1998. Evaluation of installation stock based (*R*, *Q*)-policies for two-level inventory systems with poisson demand. Operations Research 46 (3), S135–S145.

Axsater, S., 2000. Exact analysis of continuous review (*R*, *Q*) policies in two-echelon inventory systems with compound Poisson demand. Operations Research 48 (5), 686–696.

Banks, J., Spoerer, J.P., 1986. Inventory policy for the continuous review case: a simulation approach. Annals of the Society of Logistics Engineering 1 (1), 51–65.

Cohen, M.A., Kleindorfer, P., Lee, H.L., 1986. Optimal stocking policies for low usage items in multi-echelon inventory system. Naval Research Logistics Quarterly 33, 17–38.

Cohen, M., Kamesam, P., Kleindorfer, P., Lee, H.L., Tekerian, A., 1990. Optimizer: IBM's multi-echelon inventory system for managing service logistics. Interfaces 20 (1), 65–82.

Deuermeyer, B., Schwarz, L.B., 1981. A model for the analysis of system service level in warehouse/retailer distribution systems: the identical retailer case. In: Schwarz, L.B. (Ed.), Studies in Management Sciences: Multi-LevelProduction/Inventory Control Systems, Vol. 16. North-Holland, Amsterdam, pp. 163–193.

Forsberg, R., 1996. Exact evaluation of (*R*, *Q*)-policies for two-level inventory systems with poisson demand. European Journal of Operational Research 96, 130–138.

Forsberg, R., 1997. Evaluation of (*R*, *Q*)-policies for two-level inventory systems with generally distributed customer inter-arrival times. European Journal of Operational Research 99, 4001–4411.

Fortuin, L., 1980. Five popular probability density functions: a comparison in the field of stock-control models. Journal of the Operational Research Society 31, 937–942.

Fotopoulos, S., Wang, M.C., Rao, S.S., 1988. Safety stock determination with correlated demands and arbitrary lead times. European Journal of Operational Research 35, 172–181.

Graves, S.C., 1999. A single-item inventory model for a non-stationary demand process. Manufacturing and Service Operation Management 1 (1), 50–61.

Hopp, W.J., Spearman, M.L., 2000. Factory physics: foundations of manufacturing management, 2nd Edition. Irwin/MacGraw-Hill, New York.

Lagodimos, A.G., De Kok, A.G., Verrijdt, J.H.C.M., 1995. The robustness of multi-echelon service models under autocorrelated demands. Journal of the Operational Research Society 46 (1), 92–103.

Lau, H.S., Zaki, A., 1982. The sensitivity of inventory decisions to the shape of lead time demand distribution. IIE Transactions 14 (4), 265–271.

Law, A.M., Kelton, W.D., 2000. Simulation Modeling and Analysis, 3rd Edition. McGraw-Hill Inc., New York.

Moinzadeh, K., Lee, H.L., 1986. Batch size and stocking levels in multi-echelon repairable systems. Management Science 32 (12), 1567–1582.

Muckstadt, J.A., 1973. A model for a multi-item, multi-echelon, multi-indenture inventory system. Management Science 20, 472–481.

Naddor, E., 1978. Sensitivity to distributions in inventory systems. Management Sciences 24 (16), 1769–1772.

Needham, P.M., Evers, P.T., 1998. The influence of individual cost factors on the use of emergency transshipments. Transportation Research Part E: Logistics and Review 34 (2), 149–160.

Sherbrooke, C.C., 1968. METRIC: a multi-echelon technique for recoverable item control. Operations Research 16, 122–141.

Silver, E.A, Pyke, D.F., Peterson, R., 1998. Inventory management and production planning and scheduling, 3rd Edition. Wiley, New York.

Svoronos, A., Zipkin, P., 1988. Estimating the performance of multi-level inventory systems. Operations Research 36 (1), 57–72.

Tee, Y.S., Rossetti, M.D., 2001. Using simulation to evaluate a continuous review (*R*, *Q*) two-echelon inventory model, Proceedings of the Sixth Annual International Conference on Industrial Engineering—Theory, Application, and Practice, San Francisco, CA, November 18–20.

Treharne, J.T., Sox, C.R., 1999. Adaptive inventory control for non-stationary random demand (www.eng.auburn.edu/~crsox/NSFAdaptInvy/start.html), under revision for Management Science.

Tyworth, J.E., O'Neill, L., 1997. Robustness of the normal approximation of lead-time demand in a distribution setting. Naval Research Logistics Quarterly 44 (2), 165–186.

Zipkin, P., 2000. Foundations of inventory management. McGraw-Hill, New York.