

Establishing inventory policies through bootstrap estimations

Applicant: Luis Pérez - Advisor: Marco A. De Luna

November 2012

This document, written on November 2020, serves as a summary of the research thesis conducted by the author in order to obtain the degree of *Master in Science with a specialty in Quality Systems and Productivity*. With the absolute approval of the dissertation evaluation committee, this title was granted to the applicant by the Monterrey Institute of Technology and Higher Education – ITESM– in November 2012.

Abstract

One common assumption found in several inventory management models is that the lead time demand follows a normal probability distribution. If such models are used in spite of the normality premise being false, the resultant policies might negatively impact the system performance by having excessive inventory or achieving a low customer service level. Both of these results are undesirable since in the case of the former the holding cost would rise, whereas in the latter the shortage cost would grow.

This research focuses on evaluating the effectiveness of a proposed distribution free algorithm that determines the inventory target level (S) in a periodic review model (S, T). The algorithm relies heavily on the statistical tool known as bootstrap although several additional techniques are involved too.

The data used for this investigation corresponds to the actual observed behavior of 31 items sold by an independent electrical and lighting retailer. Through Dynamic Simulation a benchmark between the resulting policies from both the traditional (S, T) model and the proposed algorithm is performed; the corresponding performances are measured in terms of the ability to reach the desired customer service level along with the associated cost.

The study reveals that despite the fact that none of these approaches can guarantee that the desired service level will be achieved, the proposed algorithm has a higher chance of doing it. Moreover, the costs linked to the policies determined by the presented algorithm turned out to be lower in most cases.

keywords: *Inventory policies, (S, T) , bootstrap, distribution-free, dynamic simulation, retail distribution*

Introduction and background

According to Chopra and Meindl (2007) successful supply chain management requires taking decisions regarding the flow of information, products and resources. Particularly in the planning phase, companies define the policies that rule their operations in the short term. Among them, the inventory policies stand out for their relevance.

Furthermore, the financial success of a company depends to a great extent on the adequacy of the inventory policies to satisfy the demand it faces. If the demand is not met sales will be lost. In the other hand, an excessive inventory leads to a waste of resources.

One of the most important models to manage inventories is the periodic review model, also known as (S, T) model. It controls the inventory by issuing purchase orders –aka. PO– at intervals with a predefined length of T . Once the interval has elapsed, the positional inventory reaches a given inventory target level (S) by issuing a PO.

Waller et al. (2008) point out that even though the adoption of the (S, T) model is quite frequent, its study has been scarce. In their review, only few articles –Blumenfeld et al. (1984), Hsu and El-Najdawi (1991), Ballou (2005) and Sezen (2006)– use the periodic review system either as one of several inventory control policies or the only one.

Unfortunately, a wide spread *assumption* found in the relevant literature is that the lead time demand properly fits a normal probability distribution. In consequence, this assumption underlies in the "run of the mill" formulas generally used by professionals and coded into commercial ERP software packages. Obviously, if such formulas are used despite the lead time demand being non-normal, the resulting policies will most likely generate either lost sales or an inventory excess; both outcomes are undesirable since they negatively impact the financial performance of the company.

The periodic review model (S, T)

In the (S, T) model, the positional inventory, composed by the on-hand inventory and the in-transit inventory, is calculated at intervals with a fixed length, T . If the positional inventory turns out to be below its target level (S), a purchase order for a quantity that covers up for the difference (Q) is issued.

$$I_{\text{positional}} = I_{\text{on-hand}} + I_{\text{in-transit}} \quad (1)$$

$$Q = S - I_{\text{positional}} \quad (2)$$

The common application of the (S, T) model consists of determining the positional inventory target level $-S-$ given a previously defined review interval length T . The frequently used notation is shown in table 1.

D	Demand per time unit.
σ	Demand standard deviation per time unit.
τ	Lead time; Assumed deterministic.
T	Revision period length.
S	Positional inventory target level.

Table 1: The (S, T) model notation

The *traditional* (S, T) model

As discussed, the *traditional* model works with an stochastic revision and lead time demand (RLTD) which is *assumed* to be normal. Furthermore, another quite simplistic approach of the traditional model is to treat the lead time as a deterministic variable. With this in consideration, a typical behavior of the on-hand inventory through time is illustrated on figure 1.

For the traditional model the notation consists of what was stated in table 1 in addition to the content of table 2.

\bar{D}	Average demand per time unit.
$\bar{D}_{T+\tau}$	Average revision and lead time demand.
$\sigma_{T+\tau}$	Revision and lead time demand std. deviation.
$z_{(1-\alpha)}$	$1 - \alpha$ percentile value of the standard normal distribution probability that matches the desired in-stock probability, i.e., target service level.

Table 2: Complimentary notation for the traditional (S, T) model

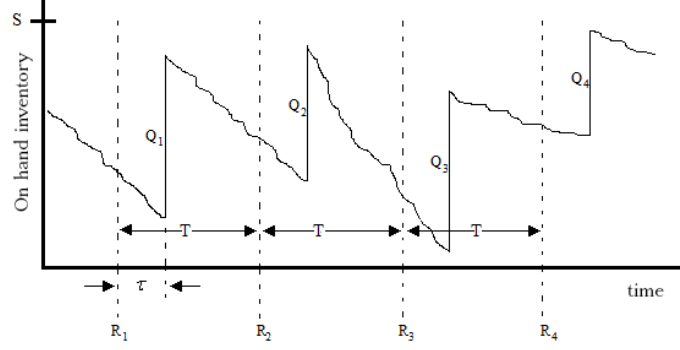


Figure 1: Example of the inventory behavior in a traditional (S, T) model

The mean and the standard deviation of the RLTD are given by equations 3 and 4 respectively.

$$\overline{D}_{T+\tau} = \overline{D}(T + \tau) \quad (3)$$

$$\sigma_{T+\tau} = \sigma\sqrt{T + \tau} \quad (4)$$

Due to the fact that the RLTD is treated as normal, the positional inventory target level (S) is defined as follows:

$$S = \overline{D}_{T+\tau} + z_{(1-\alpha)}\sigma_{T+\tau} \quad (5)$$

Figure 2 shows an example of a KDE plot of the RLTD along with the S estimation location over the $D_{T+\tau}$ axis for this scenario. In order to ease the grasp of the algorithm that will be presented next, it is extremely important to realize that as shown in this figure, the positional inventory target level S is nothing else than the $1 - \alpha$ percentile value of the RLTD. In this context, $1 - \alpha$ represents the targeted in-stock probability, i.e. desired service level.

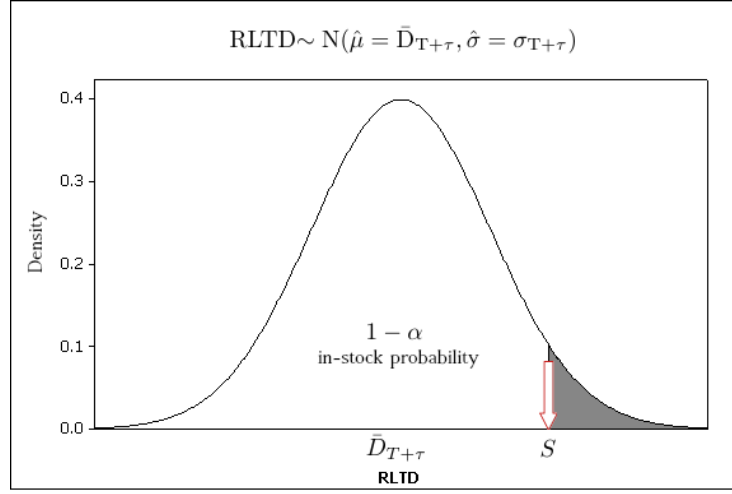


Figure 2: KDE plot of the $RLTD$ in the traditional (S, T) model

More objections to the traditional (S, T) model

Besides the main objection to the assumption of normality there are a couple more aspects of the traditional (S, T) model that are highly questionable by those who have had "field" experience in determining inventory policies.

- τ treated as deterministic. In the day-to-day reality lead times can vary greatly; delays from the supplier take place up to the point of τ being several times higher than T making the behavior of the system something similar to what is shown in figure 3.
- Regardless of we consider $\tau < T$ or $\tau > T$, undoubtedly τ and T overlap each other. In light of this, the logic behind equation 5 is quite dubious. Therefore to emphasize this disagreement, from this point D_ω will be used to represent the demand that takes place during the overlapped lead time and revision period, replacing $D_{T+\tau}$.

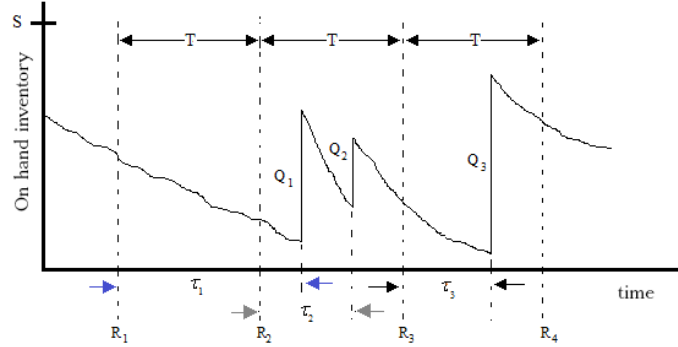


Figure 3: Example of the inventory behavior in a realistic (S, T) model

Overview of the relevant statistical tools

The algorithm to be presented in the next section relies in several statistical tools which quick revisions are deemed as pertinent in order to ease the reading of this document. Namely:

- The bootstrap - Efron (1979).
- The jackknife - Quenouille (1949).
- Asymmetry significance testing - Tabachnick and Fidell (2000).
- Bootstrap BC_α confidence intervals - Efron and Tibshirani (1994).
- Data transformation for robust bootstrap estimation - Singh et al. (1998).

The bootstrap

The bootstrap methods –Efron (1979)– depend on the notion of a *bootstrap sample*. Let \hat{F} be an empirical distribution of the random variable X , putting a probability of $1/n$ on each of the observed values x_i , $i = 1, 2, \dots, n$. A bootstrap sample is defined to be a random sample of size n drawn from \hat{F} , say

$$\hat{F} \rightarrow [x_1, x_2, \dots, x_n] \quad (6)$$

$$\vec{x}_b = [x_{1_b}, x_{2_b}, \dots, x_{n_b}] \quad (7)$$

The b sub index indicates that \vec{x}_b is not the actual data set \vec{x} , but rather a randomized, or re-sampled, version of it. In other words, the bootstrap data points are a random sample of size n *with replacement* from the original observed values x_1, x_2, \dots, x_n . Thus we might have $x_{1_b} = x_7, x_{2_b} = x_3, \dots, x_{n_b} = x_1$. The bootstrap data set $[x_{1_b}, x_{2_b}, \dots, x_{n_b}]$ consists of members of the original data set $[x_1, x_2, \dots, x_n]$ some appearing zero times, some appearing twice, etc.

Corresponding to a bootstrap sample \vec{x} , a *bootstrap replica* is defined as:

$$\hat{\theta}_b = s(\vec{x}_b) \quad (8)$$

Where $\hat{\theta}_b$ is the results of applying any given function s to the data set $[x_{1_b}, x_{2_b}, \dots, x_{n_b}]$ as if it was applied to a *regular* sample of the random variable X , in order to estimate a parameter of interest θ . The notation $\hat{\theta}$ indicates that we are talking about an *estimation* of the parameter θ , not its true value, whereas $\hat{\theta}_b$ is used to represent an estimation of the parameter of interest calculated from a bootstrap sample.

A bootstrap algorithm works by drawing many independent samples, and evaluating the corresponding bootstrap replicas which results are stored in a vector e.g. $\vec{\hat{\theta}}_b$. From there a punctual estimation of the parameter of interest as well as the standard error –given by the sample standard deviation– can be calculated as:

$$\hat{\theta}_{b(\cdot)} = \sum_{\hat{\theta}_b \in \vec{\hat{\theta}}_b} \hat{\theta}_b \cdot (1/\#\vec{\hat{\theta}}_b) \quad (9)$$

$$\hat{s}e_{\hat{\theta}_b} = \left[\sum_{\hat{\theta}_b \in \vec{\hat{\theta}}_b} \left(\hat{\theta}_b - \hat{\theta}_{b(\cdot)} \right)^2 / \left(\#\vec{\hat{\theta}}_b - 1 \right) \right]^{1/2} \quad (10)$$

Let B represent the number of bootstrap replicas. As $B \rightarrow \infty$, $\hat{s}e_{\hat{\theta}_b}$ approaches $\hat{s}e_{\hat{\theta}_F}$. In consequence, the number of bootstrap replicas – B – that are enough to properly estimate the standard error is between 25 to 200.

The jackknife

The jackknife is a technique for estimating both the bias and the standard error of an estimate. It precedes the bootstrap and bears close similarities to it. Suppose we have a sample $\vec{x} = [x_1, x_2, \dots, x_n]$. The jackknife focuses on re-sampling the original data set *leaving out one observation at a time*:

$$\vec{x}_j = [x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n] \quad (11)$$

for $i = 1, 2, \dots, n$, called jackknife samples. The i th jackknife sample consists of the original data set with the i th observation removed. Let

$$\hat{\theta}_{j(i)} = s(\vec{x}_{j(i)}) \quad (12)$$

be the i th jackknife replication of $\hat{\theta}$. The jackknife estimate of bias is defined by

$$\hat{\text{bias}}_{\text{jack}} = (n - 1)(\hat{\theta}_{j(\cdot)} - \hat{\theta}) \quad (13)$$

where

$$\hat{\theta}_{j(\cdot)} = \sum_{i=1}^n \hat{\theta}_{j(i)} / n \quad (14)$$

Bootstrap BC_a confidence intervals

Efron and Tibshirani (1994) remark that one of the principal goals of the bootstrap is to produce good confidence intervals. In this context, "good" means that the bootstrap intervals should closely match exact confidence intervals in those special situations where statistical theory yields an exact answer, and should give dependably accurate coverage probabilities in all situations.

Among the several methods available to construct confidence intervals by making use of the bootstrap, BC_a – *bias corrected and accelerated* – intervals are the ones that come closest to the before mentioned criteria of goodness; The downside of the BC_a confidence intervals is that they are also the most computationally expensive to produce as they involve the use of the jackknife.

The bias correction coefficient, z_0 , is calculated by

$$\hat{z}_0 = \Phi^{-1} \left[\frac{\# \left(\left\{ \hat{\theta}_b \in \vec{\hat{\theta}}_b \right\} < \hat{\theta} \right)}{\#\vec{\hat{\theta}}_b} \right] \quad (15)$$

where Φ^{-1} is the standard normal inverse cumulative distribution function. For example $\Phi^{-1}(0.95) = 1.645$.

The acceleration coefficient \hat{a} refers to the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter value of θ , *measured on a normalized scale*. It is calculated with the formula:

$$\hat{a} = \frac{\sum_{i=1}^n \left(\hat{\theta}_{j(\cdot)} - \hat{\theta}_{j(i)} \right)^3}{6 \left\{ \sum_{i=1}^n \left(\hat{\theta}_{j(\cdot)} - \hat{\theta}_{j(i)} \right)^2 \right\}^{3/2}} \quad (16)$$

Let α represent the type I error probability. The BC_a interval of intended coverage $1 - 2\alpha$ is given by

$$\text{BC}_a : \left[\hat{\theta}_{\text{bLo}}, \hat{\theta}_{\text{bUp}} \right] = \left[\text{percentile}_{\alpha_1} \left(\vec{\hat{\theta}}_b \right), \text{percentile}_{\alpha_2} \left(\vec{\hat{\theta}}_b \right) \right] \quad (17)$$

where

$$\alpha_1 = \Phi \left[\hat{z}_0 + \frac{\hat{z}_0 + z_{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z_{(\alpha)})} \right] \quad (18)$$

$$\alpha_2 = \Phi \left[\hat{z}_0 + \frac{\hat{z}_0 + z_{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z_{(1-\alpha)})} \right] \quad (19)$$

Here Φ represents the standard normal cumulative distribution function whereas $z_{(\alpha)}$ is the 100 α th percentile value of the standard normal probability distribution. For example $z_{(0.95)} = 1.645$ and $\Phi(1.645) = 0.95$.

Asymmetry significance testing

When a random variable follows a normal probability distribution its asymmetry and kurtosis values are equal to zero. According to Tabachnick and Fidell (2000) there are asymmetry significance tests that try the actual value against the null hypothesis of zero. This is

$$H_0 : S = 0 \quad (20)$$

$$H_1 : S \neq 0 \quad (21)$$

The standard error of the asymmetry can be estimated by

$$\hat{se}_S = \sqrt{6/n} \quad (22)$$

where n is the quantity of observations in the sample. The asymmetry coefficient calculated from the sample $-S-$ is then compared to the value of 0 proper of a standard normal distribution. This will enable us to calculate z by using

$$z = \frac{S - 0}{\hat{se}_S} \quad (23)$$

which in turn will allow us to determine the p -value required to perform the hypothesis test as

$$p\text{-value} = 2[1 - \Phi(z)] \quad (24)$$

where as usual, Φ represents the cumulative function of the standard normal distribution. If with the α significance level H_0 is rejected, i.e. $p\text{-value} < \alpha$, we could conclude that statistically the observed S coefficient is different than zero and therefore the asymmetry in the sample is significant.

Data transformation for robust bootstrap estimation

Singh et al. (1998) state that in the robust bootstrap estimation, extreme values have little to no influence in the obtained estimator. In consequence, some of the values at the extremes of \vec{x}_s –the *sorted* version of the sample data set \vec{x} – could be modified to further increase the robustness of bootstrap percentiles.

In a data set \vec{x}_s with n observations the integer l is calculated by

$$l = \lfloor n\beta \rfloor \quad (25)$$

where the coefficient β , with a value between 0 and $1/2$, represents the β –th fraction of the data that will be transformed according to the logic of figure 4.

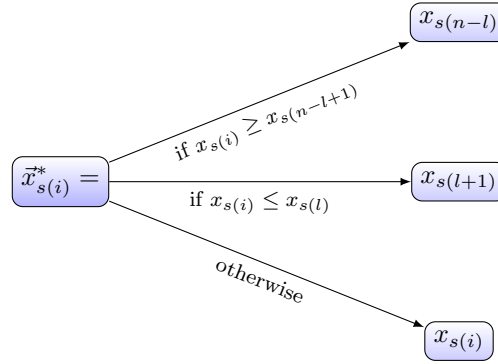


Figure 4: *winsorization* data transformation logic
Singh et al. (1998)

Here $\vec{x}_{s(i)}^*$ represents the transformed –*winsorized*– data set from which the bootstrap samples will be generated. Singh et al. (1998) sustain that despite the fact any $\beta < 1/2$ value is theoretically valid, in practice β should be set significantly below that value, e.g. $\beta \leq 1/4$.

A new proposal for the estimation of S

Assumptions and required information

The algorithm takes a more realistic approach compared to the traditional model by assuming that the available information about the RLTD, D_ω , is in fact accessible through *real-life observations* of two independent random variables:

- Demand per time unit, D .
- Lead time, τ .

Additionally, the following parameters must be previously defined:

- Review interval length, T .
- Desired service level, $1 - \alpha$, i.e. the in-stock probability.

During the research the hardest obstacle to overcome was to obtain the *actual demand* information for the studied items. The reason for this is that the vast majority of the companies do keep a record of their sales figures, but in the case of the demand its data is rarely available.

It is worth recalling that $1 - \alpha$ corresponds to the desired in-stock probability and that consequently the $1 - \alpha$ percentile of D_ω determines the S value in a periodic review system. Accordingly, θ will be used to represent our parameter of interest i.e. $\theta = \text{percentile}_{1-\alpha}(D_\omega)$.

The pseudo code for the algorithm that estimates the positional inventory target level (S), is divided across 4 main blocks:

1. Data asymmetry testing and transformation.
2. Generation of D_ω sample values.
3. Evaluation of $\hat{\theta}$ bootstrap replicas.
4. Construction of BC_α confidence intervals for θ .

Algorithm Block 1: data asymmetry testing and transformation

A common problem while working with demand and lead time data is the presence of extreme values. Such observations might cause that the empirical distribution of D_ω to end up with a long right tail; As a result its estimated $1 - \alpha$ percentile value would be significantly higher, leading to a bigger positional inventory target level and thus a larger holding cost.

Since the presence of extreme values would most likely result in a higher asymmetry, a hypothesis test to evaluate whether or not the asymmetry turns out to be significant is performed as proposed by Tabachnick and Fidell (2000). In the case that the null hypothesis is rejected, i.e. $H_0 : S = 0$, a portion of the observations at both ends will be transformed using the method suggested by Singh et al. (1998).

Another situation that could arise when working with real data of D is the presence of zero –no demand– or negative values –returns surpassing the demand– among the observations. Johansen and Hill (2000) make a justification to drop the negative values; As for the zero values, it was decided to also drop them in order to aim for a resulting higher service level.

1. Create the vectors –1-D arrays– $\vec{\tau}$ and \vec{d} from the available τ and D observations.
2. Perform an asymmetry test on the $\vec{\tau}$ and \vec{d} values.
 - If H_0 is rejected, transform the data of the corresponding vector.

Algorithm Block 2: generation of D_w sample values

Up to this point, the available information of the RLTD is available on $\vec{\tau}$ and \vec{d} values. By making use of the bootstrap, $\vec{\tau}$ and \vec{d} observations will be mixed to generate D_w sample values.

3. Initialize \vec{d}_w as an empty vector,

$$\vec{d}_w = \emptyset \quad (26)$$

4. Determine the bootstrap sample size, n , that will be used as follows:

$$n = \left\lceil \frac{\#\vec{d} + \#\vec{\tau}}{2} \right\rceil \quad (27)$$

5. While $\#\vec{d}_w < 300$:
 - (a) Generate a bootstrap sample of size n from $\vec{\tau}$, $\vec{\tau}_b$
 - (b) Generate a vector of size n , \vec{t} where $\forall t \in \vec{t}, t = T$
 - (c) $\vec{m}_b = \max(\vec{\tau}_b, \vec{t})$
 - (d) Generate a bootstrap sample of size n from \vec{d} , \vec{d}_b
 - (e) $\vec{d}_{wb} = \vec{m}_b \odot \vec{d}_b$
 - (f) Add the values of \vec{d}_{wb} to \vec{d}_w
 - (g) $\vec{\tau}_b = \vec{m}_b = \vec{d}_b = \vec{d}_{wb} = \emptyset$

The condition established in the fifth step determines that the number of elements in \vec{d}_w should be at least 300. This sample size for D_w certainly seems reasonable if the rule of thumb $B = 200$ set by Efron and Tibshirani (1994) to estimate standard errors is considered as a reference.

Algorithm Block 3: evaluation of $\hat{\theta}$ bootstrap replicas

Now that sample values of D_ω are available, the bootstrap will be used once again in order to make estimations of the parameter of interest θ , which is the $1 - \alpha$ percentile value of D_ω .

6. Initialize $\vec{\hat{\theta}}_b$ as an empty vector,

$$\vec{\hat{\theta}}_b = \emptyset \quad (28)$$

7. While $\#\vec{\hat{\theta}}_b < 2000$:

- (a) Take a bootstrap sample from $\vec{d}_\omega, \vec{d}_{w_b}$
- (b) Evaluate the bootstrap replica $\hat{\theta}_b$ as follows:

$$\hat{\theta}_b = \text{percentile}_{1-\alpha}(\vec{d}_{w_b}) \quad (29)$$

- (c) Add the $\hat{\theta}_b$ value to $\vec{\hat{\theta}}_b$
- (d) $\vec{d}_{w_b} = \emptyset$

A punctual estimation of $\hat{\theta}$, $\hat{\theta}_{b(\cdot)}$, could be simply determined by averaging the values of $\vec{\hat{\theta}}_b$. This is:

$$\hat{\theta}_{b(\cdot)} = \sum_{\hat{\theta}_b \in \vec{\hat{\theta}}_b} \hat{\theta}_b \cdot \left(1/\#\vec{\hat{\theta}}_b\right) \quad (30)$$

Since $\hat{\theta}_{b(\cdot)}$ is an estimation of the $1 - \alpha$ percentile value of D_ω , we could obviously just set it as S in the system and be done with our task. Nonetheless, for an enhanced trust in the result, a confidence interval for θ can be constructed. This idea will be further discussed in the next subsection.

Algorithm Block 4: Construction of BC_a confidence intervals for θ

According to Devore (2015) since a punctual estimation is just a number, it does not provide any information of its precision nor reliability. Due to the sampling variability the case of a $\hat{\theta} = \theta$ never virtually occurs. Moreover, a punctual estimation does not provide any information about the closeness of $\hat{\theta}$ to θ . Such inconveniences can be overcome by reporting estimation intervals, i.e. confidence intervals.

To conclude the algorithm, as reaching the objective service level –i.e. in stock probability– has been prioritized over cost reduction, the positional inventory target level, S , is set as the $\hat{\theta}$ BC _{α} confidence interval upper limit value.

12. Compute all jackknife replicas of $\hat{\theta}$, $\hat{\theta}_{j(i)}$, storing all their values in $\vec{\hat{\theta}}_j$
13. Calculate the jackknife punctual estimator of θ , $\hat{\theta}_{j(\cdot)}$
14. Determine the acceleration \hat{a} and bias correction \hat{z}_0 coefficients
15. Assess the α_1 and α_2 cumulative probabilities
16. Evaluate the α_1 and α_2 percentile values of \vec{d}_{ω_b} to respectively determine the lower and upper limits of the $\hat{\theta}_b$ BC _{α} confidence interval. In other terms $\hat{\theta}_{bLo}$ and $\hat{\theta}_{bUp}$
17. Set $S = \hat{\theta}_{bUp}$

Evaluation methodology

Note: Please see the changelog section after the references.

In order to determine the effectiveness of the policies that result from applying the presented technique, a benchmark against the traditional (S, T) model was carried over with the aid of dynamic simulation. For all of the items in the sample, 50 simulation runs per method, each one with a length of 350 units of time, were executed.

The comparison of the performance was done in terms of the ability to reach the desired customer service level along with the associated cost. The latter is integrated by the costs of ordering, holding the inventory and the cost of shortages. An scenario of lost sales as response from the customer was assumed whereas the fixed shortage cost per unit was determined to be of the same magnitude of the lost profit. The ordering cost was set at \$30 USD. Additionally, the yearly interest rate used to determine the holding cost was 40%.

An EDA of the simulation results was realized and as a consequence they were suspected to be non-normal. Thus the non-parametric Mann-Whitney U test –Mann and Whitney (1947)– was deemed as appropriate to perform the related hypothesis tests. In this test, also known as Wilcoxon rank-sum, randomly selected values from two independent samples are dealt with. The null hypotheses states that the probabilities of a random value taken from any of the samples being greater than its counterpart is the same than the counterpart being greater than the random taken value, i.e.:

$$H_0 : P [\vec{x}_{(i)} > \vec{y}_{(j)}] = P [\vec{y}_{(j)} > \vec{x}_{(i)}] \quad (31)$$

here i and j are independent random integers between 1 and the samples size, n . With this in consideration and making a general use of G and H to denote the results from the presented and traditional methods respectively, table 3 indicates the *raw* alternate hypotheses relevant to the benchmark:

Performance indicator	Alternate hypothesis
Service level	$H_1 : P [\vec{g}_{SL(i)} > (1 - \alpha)] > P [(1 - \alpha) > \vec{g}_{SL(i)}]$
Service level	$H_1 : P [\vec{h}_{SL(j)} > (1 - \alpha)] > P [(1 - \alpha) > \vec{h}_{SL(j)}]$
Service level	$H_1 : P [\vec{g}_{SL(i)} > \vec{h}_{SL(j)}] > P [\vec{h}_{SL(j)} > \vec{g}_{SL(i)}]$
Total cost	$H_1 : P [\vec{h}_{K(j)} > \vec{g}_{K(i)}] > P [\vec{g}_{K(i)} > \vec{h}_{K(j)}]$

Table 3: benchmark alternate hypotheses

The tests were performed with pingouin –Vallat (2018)– which by default correct for ties and by default use a continuity correction; the rank biserial correlation –Kerby (2014)– is the difference between the proportion of favorable evidence minus the proportion of unfavorable evidence.

The common language effect size, a concept introduced by McGraw and Wong (1992), is the proportion of pairs where x is higher than y . Pingouin uses a brute-force version of the formula given by Vargha and Delaney (2000):

$$CL = P [\vec{x}_{(i)} > \vec{y}_{(j)}] + 0.5 \times P [\vec{x}_{(i)} = \vec{y}_{(j)}] \quad (32)$$

The advantage of this method is twofold: The brute-force approach pairs each observation of x to its y counterpart, and therefore does not require normally distributed data. Additionally, the formula takes ties into account and therefore works with ordinal data.

Conceding to the benchmark scheme of table 3, the data sets of the observed service levels are being used twice on different tests. As a result, in order to avoid spurious positives the α value should be lowered to account for the number of comparisons being performed, i.e. n . One approach is to use the Bonferroni correction –Weisstein (2004)– which sets the significance level for the entire set of comparisons to α/n . Therefore, in this scenario $\alpha_{\text{bonferroni}} = 0.05/2$.

Results overview, final thoughts and conclusions

The portions of the cases in the study sample in which the null hypotheses were rejected in favor of the alternates –as defined in table 3– are reported in

table 4.

Alternate hypothesis	% of H_0 rejected
$H_1 : P [\vec{g}_{SL(i)} > (1 - \alpha)] > P [(1 - \alpha) > \vec{g}_{SL(i)}]$	90.32%
$H_1 : P [\vec{h}_{SL(j)} > (1 - \alpha)] > P [(1 - \alpha) > \vec{h}_{SL(j)}]$	70.97%
$H_1 : P [\vec{g}_{SL(i)} > \vec{h}_{SL(j)}] > P [\vec{h}_{SL(j)} > \vec{g}_{SL(i)}]$	70.97%
$H_1 : P [\vec{h}_{K(j)} > \vec{g}_{K(i)}] > P [\vec{g}_{K(i)} > \vec{h}_{K(j)}]$	74.19%

Table 4: benchmark alternate hypotheses

With a 19.35% absolute higher chance of achieving the desired customer service level, its clearly shown that the proposed distribution-free algorithm represents a considerable improvement over the traditional method which assumes normality on the RLTD.

By using the presented technique, the probability of getting a lesser total cost than is counterpart is 74.19%. Moreover, the cases in which a greater in-stock probability and a lower cost were attained *simultaneously* –i.e. the last two hypotheses from table 3 rejected– accounted for 70.97% of the total. As a side note, it is worth pointing out that the cost benefits may vary greatly according to the parameters used to determine the total cost.

From the experience of the author, it is very likely that for any company who wishes to implement this algorithm, the hardest challenge to face would be the availability of the lead time and, especially, demand information. Without reliable data collected systematically, any application effort would end up being utterly worthless. For many companies this would mean a re-design –at least partial– of their commercial processes, since *demand* records are rarely available.

Given the positive results of this research, it is easy to realize how important it is to not make statistical assumptions lightly –e.g. demand following a normal probability distribution, fixed lead time– and rather be analytical when implementing any kind of solution that has a statistical basis. Although for the statistical-untrained eye the presented distribution-free algorithm might seem complex, the potential benefits make it undoubtedly attractive.

References

- Ballou, R. H. (2005). Expressing inventory control policy in the turnover curve. *Journal of Business Logistics*, 26(2):143–164.

- Blumenfeld, D. E., Hall, R. W., and Jordan, W. C. (1984). Trade-off between freight expediting and safety stock inventory costs. Technical report.
- Chopra, S. and Meindl, P. (2007). *Supply Chain Management: Strategy, Planning, and Operation*. Pearson education. Pearson Prentice Hall.
- Devore, J. L. (2015). *Probability and Statistics for Engineering and the Sciences*. Cengage Learning.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. the annals of statistics, 7 (1): 1–26. URL <http://www.jstor.org/stable/2958830>.
- Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Hsu, J. I. and El-Najdawi, M. (1991). Integrating safety stock and lot-sizing policies for multi-stage inventory systems under uncertainty. *Journal of Business Logistics*, 12(2):221.
- Johansen, S. G. and Hill, R. M. (2000). The (r, q) control of a periodic-review inventory system with continuous demand and lost sales. *International Journal of Production Economics*, 68(3):279–286.
- Kerby, D. S. (2014). The simple difference formula: An approach to teaching nonparametric correlation. *Comprehensive Psychology*, 3:11–IT.
- Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60.
- McGraw, K. O. and Wong, S. P. (1992). A common language effect size statistic. *Psychological bulletin*, 111(2):361.
- Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):68–84.
- Sezen, B. (2006). Changes in performance under various lengths of review periods in a periodic review inventory control system with lost sales. *International Journal of Physical Distribution & Logistics Management*.
- Singh, K. et al. (1998). Breakdown theory for bootstrap quantiles. *The annals of Statistics*, 26(5):1719–1732.
- Tabachnick, B. G. and Fidell, L. S. (2000). Using multivariate statistics (vol. 1). northridge.
- Vallat, R. (2018). Pingouin: statistics in python. *The Journal of Open Source Software*, 3(31):1026.
- Vargha, A. and Delaney, H. D. (2000). A critique and improvement of the cl common language effect size statistics of mcgraw and wong. *Journal of Educational and Behavioral Statistics*, 25(2):101–132.

Waller, M., Williams, B. D., and Tokar, T. (2008). A review of inventory management research in major logistics journals. *The International Journal of Logistics Management*.

Weisstein, E. W. (2004). Bonferroni correction. <https://mathworld.wolfram.com/>.

Changelog

2020 November

- Notation was completely overhauled mainly to make it consistent across the whole document and ease the reading.
- When preparing this summary it was decided to re-do the statistical analysis of the simulation results. Under fresh eyes, the original one seemed unnecessarily complex. Hence the reference to 3 publications with a date passed the writing of the original thesis. *The conclusions remained the same.*