Estimating the Cumulative Distribution Function of Lead-Time Demand using Bootstrapping With and Without Replacement

John E. Boylan[1,*]   M. Zied Babai[2]

[1]Centre for Marketing Analytics and Forecasting, Lancaster University, United Kingdom

[2]Centre of Excellence in Supply Chain - CESIT, Kedge Business School, Bordeaux, France

*Corresponding author: j.boylan@lancaster.ac.uk

Abstract

Forecasting of the cumulative distribution function (CDF) of demand over lead time is a standard requirement for effective inventory replenishment. In practice, while the demand for some items conforms to standard probability distributions, the demand for others does not, thus making it challenging to estimate the CDF of lead-time demand. Distribution-free methods have been proposed, including resampling of demand from previous individual periods of the demand history, often referred to as bootstrapping in the inventory forecasting literature. There has been a lack of theoretical research on this form of resampling.

In this paper, we analyze the bias and variance of CDF estimates obtained by resampling, both with and without replacement. Counterintuitively, we find that the 'with replacement' approach does not always dominate 'without replacement' in terms of mean square error of CDF estimates. Closed-form expressions are given for the components of Mean Square Error, with and without replacement. For shorter lead times, of two or three periods, these may be used directly to identify series that may benefit from resampling without replacement. Inventory performance implications are evaluated on simulated and empirical data. It is found that marked differences may arise between 'with replacement' and 'without replacement' bootstrapping approaches. The latter method can be more beneficial for lower target Cycle Service Levels, longer lead times and shorter demand histories.

Keywords: *Forecasting, resampling, bootstrapping, demand distribution, inventory*

1 Introduction

This paper is concerned with the statistical properties of resampling approaches that have been recommended for inventory replenishment systems, and their implications for inventory performance. In this introduction, the background of the research is outlined, a brief literature review is presented, and our findings are summarized.

1.1 Background

Inventory replenishment systems often require, at the level of the individual stock keeping unit (SKU), forecasting of the cumulative distribution function (CDF) of demand over the lead time (Siddiqui et al. 2022). The characterization of the cumulative distribution function of lead-time demand is often a challenging problem in practice. Many probability distributions have been suggested, with the normal and gamma being popular choices for faster-moving demand, and the Poisson and compound Poisson distributions finding favour for slower-moving demand. Amongst the compound Poisson distributions, the negative binomial (Poisson-logarithmic) and stuttering Poisson (Poisson-geometric) are often recommended (see, for example, Axsäter (2015)). The characterization of the demand of all items by compound Poisson distributions can become inaccurate, especially when the probability mass functions become multi-modal, as often observed in industry.

Syntetos et al. (2013) undertook an extensive empirical analysis of the goodness of fit of the Poisson, normal, gamma, negative binomial, and stuttering Poisson distributions to discrete demand patterns, examining 5000 SKUs from the Royal Air Force (RAF) and just over 3000 series from the electronics sector. They found that all distributions failed to fit lead-time demand for at least 10% of the electronics demand series, and for more than 25% of the RAF series, according to a Kolmogorov-Smirnov (KS) goodness-of-fit test at the 5% level. These findings were confirmed by Turrini and Meissner (2019), who re-analyzed the RAF dataset, using the same distributions as Syntetos et al. (2013) but employing a modified KS test that puts more emphasis on the upper tail of the distribution, reflecting the importance of this tail in inventory management. They found that all distributions failed to achieve a 'strong fit' (at the 5% level) to lead-time demand for at least 50% of series. They also noted that a Kolmogorov-Smirnov test is conservative for discrete data, meaning that the results may be an overestimate of the percentages of series well fitted by the distributions tested.

For items with non-standard demand distributions (and standard distributions alike) distribution-free approaches may be used for the forecasting of the cumulative distribution function of demand over the lead time. One such approach is based on resampling, with or without replacement, previous demand observations over individual periods. The resampling approach, often referred to as bootstrapping in the inventory forecasting literature, has been adopted, with modifications (Willemain and Smart (2001)), in commercial demand forecasting software, in current use by a wide range of organizations. Nevertheless, its properties, with specific reference to the cumulative distribution function, have not been well studied. In this paper, we endeavour to fill this gap, by establishing statistical properties of cumulative distribution function estimates from resampling.

1.2 Literature Review

Akcay et al. (2011) categorized research in statistical estimation of demand distributions into three streams, according to the available information on the distribution and its parameters. In the first stream, the form of the demand distribution is known but its parameters are unspecified. In the second stream, partial information about the demand distribution is available. For example, Saghafian and Tomlin (2016) considered the case where a set of possible demand distributions is known and there are bounds on the expected value of demand and the tail probabilities. In the third stream of research, the demand is modelled by an empirical distribution function of the historical demand data.

The first stream of research is the most well developed, but is unhelpful when demand distributions are not 'well behaved', for example if they are multi-modal. In the second stream of research, replenishment quantitites are sought that maximize the worst-case profit (e.g. Gallego and Moon (1993)), satisfy a minimax regret criterion (e.g. Perakis and Roels (2008)), or are based on a Conditional-Value-at-Risk objective (e.g. Lee, Kim and Moon (2021)). In the third stream of research, the empirical distribution function may be estimated in different ways, depending on the assumptions about the nature of demand and the information available to the decision maker. For example, Huh et al. (2011) assumed i.i.d. discrete (censored) demand, and proposed the Kaplan-Meier estimator for constructing the empirical CDF. Other authors have focused on i.i.d. discrete (uncensored) demand, using blocking or resampling procedures.

Two commonly employed blocking procedures are non-overlapping aggregation and overlapping aggregation (Porras and Dekker (2008), Rostami-Tabar et al. (2013)). In the first approach, an empirical distribution function is calculated based on the proportions of previous blocks (of length equal to the lead time) whose total demand has not exceeded the specified amount. The second approach is similar to the first, but is based on historical blocks that overlap; for example, if the block-size is 2 and the demand history is for 24 periods, then there will be 12 non-overlapping blocks, but 23 overlapping blocks. For independent and identically distributed demand, it is well known that both of these approaches produce unbiased estimates of the CDF. Boylan and Babai (2016) gave a formula for the variance of CDF estimates from overlapping blocks, and compared it to the variance of non-overlapping blocks. They showed that overlapping blocks usually produces lower variance than non-overlapping blocks, but there are some circumstances where this result is reversed. The overlapping blocks method has been augmented by using empirical extreme value theory to model the tail of the lead-time demand distribution (Zhu et al. (2017)).

Bookbinder and Lordahl (1989) was the first paper in the inventory literature to use the bootstrapping approach (Efron (1979)) to determine the reorder point in an inventory system operating to achieve a target service level. Based on lead-time simulated data from a number of populations with varying tail shapes, they analysed the statistical and inventory cost performance of the bootstrapping approach compared to the normal distribution based approach. Wang and Rao (1992) used the bootstrapping approach to estimate the reorder point when demand follows an autoregressive process of order one (AR(1)). In both of these papers, previous lead-time demands were resampled. Fricker and Goodhart (2000) investigated the estimation of demand distributions for the Marine Expeditionary Force. They found that direct resampling of lead-time demands was not feasible. Actual lead-time periods could not be identified because historical data on the inventory position was not available. They proposed a resampling scheme that can be summarized as follows. For a fixed lead time of L periods, demands over single periods are randomly selected (with replacement). This is done L times, with the total demand representing the resampled demand over lead time. This whole process is repeated many times, thereby yielding a discrete empirical distribution function of lead-time demand, together with the reorder point/ order-up-to level corresponding to the appropriate quantile of the distribution. This approach allows for a more efficient use of the available data, for independent and identically distributed time series, as it permits the resampling of non-consecutive (and repeated) periods. It may be easily communicated to practitioners with a limited technical grounding in statistics and can be applied when data histories are short, as is common in practice. However, Fricker and Goodhart (2000) did not investigate its statistical properties.

The idea of resampling demands from single periods has been found to be particularly appealing for intermittent demand patterns which, in addition to sporadicity of demand occurrence, often show highly erratic discrete demand. Some of these demand series may exhibit serial independence but others show 'streaks' of demand occurrence (Willemain et al. (1994)). To address this issue, Willemain et al. (2004) introduced a Markov chain mechanism, allowing dependence between successive demand occurrences to be modelled, and a 'jittering' mechanism, whereby previously unobserved demands (over a single period) can be introduced into the procedure. This methodology, often referred to as WSS in the literature, has been embedded in commercial software, which has been patented in the United States (Willemain and Smart (2001)). The software has been adopted by companies in many sectors including automotive, aviation, durable goods, industrial equipment, and public transit. The 'jittering' mechanism has been critiqued (Rego and Mesquita (2015)) and sources of bias summarized (Boylan and Syntetos (2021)). Nevertheless, the fundamental resampling method, equivalent to that of Fricker and Goodhart (2000), has remained unchallenged.

Another refinement of resampling with replacement was proposed by Zhou and Viswanathan (2011) for intermittent demand series. According to this method, demand intervals (the number of periods betwen successive non-zeroes) and demand sizes (non-zeroes only) are resampled with replacement. If demand occurs in every period, then the method is equivalent to that of Fricker and Goodhart (2000). Hasni et al. (2019) conducted an empirical comparison of the accuracy of the methods advocated by Willemain et al. (2004) and Zhou and Viswanathan (2011), with neither method dominating the other over all experimental settings. However, there was no investigation of the statistical properties of the fundamental approach on which these methods are based.

In summary, resampling demands from individual periods to reconstitute lead-time demands, as recommended by Fricker and Goodhart (2000), is an attractive idea and has been taken forward by other authors. Although these methods have been evaluated empirically, the statistical properties of resampled estimates of the cumulative distribution function (CDF) of demand have not been investigated. The aim of the research presented in this paper is to make progress in that direction, by deriving some analytical results, and examining their implications for the mean square error of CDF estimates and for inventory performance.

1.3 Findings

To achieve the aim stated above, this paper makes the following contributions for discrete i.i.d. demand:

1. A bias expression is obtained for estimates of the CDF of aggregate demand over several periods, using resampling with replacement from previous individual periods. It is shown that, for an ever-increasing length of demand history, the method is asymptotically unbiased.

2. An analytical result is established for the maximum magnitude of bias for CDF estimates from resampling with replacement, for aggregation of demand over two periods.

3. Variance expressions are obtained for the estimates of the CDF of demand over several periods using resampling with and without replacement.

4. Numerical examples are given to demonstrate that the mean square error of a CDF estimate from resampling 'without replacement' can be lower than resampling 'with replacement', and to show that the reduction may be substantial.

5. Experiments are conducted on simulated and empirical data, demonstrating that 'without replacement' can generate better inventory performance than 'with replacement', especially for lower Cycle Service Level targets, longer lead times and shorter demand histories.

The first and third contributions are generally applicable to discrete i.i.d. demand for any length of aggregation in the resampling methods. The second result is specific to a length of two periods. Analytical results for the maximum magnitude of bias for aggregation of demand over three or more periods become intractable. Also, the variance expression for sampling with replacement becomes very unwieldy for longer periods of aggregation. However, by means of counter-examples, the fourth contribution is to show that 'with replacement' does not always dominate 'without replacement', in terms of mean square error, for i.i.d. demand.

The fifth contribution demonstrates that the 'without replacement' approach to this form of resampling should not be ignored. The 'with replacement' approach to the estimation of the Cumulative Distribution Function can lead to biases and greater variance in the estimates of CDFs than 'without replacement', especially when long demand histories are not available. These can translate to improvements in inventory performance by using 'without replacement' in certain situations, which are identified in this paper.

2 Bias of Resampled Estimates for the Discrete Cumulative Distribution Function

The aim of this section is to establish the bias properties of resampled estimates from previous individual periods, with and without replacement, of the cumulative distribution function (CDF) of lead-time demand. General expressions for bias are given and asymptotic properties are established, based on an ever-increasing number of historical demand observations.

2.1 Assumptions and Notations

We assume that historical demand values $d_1, d_2, ..., d_n$ are non-negative integers and are observed from an independent and identically distributed (i.i.d.) time series. These assumptions are consistent with the resampling of individual historical periods, which may be non-consecutive, as noted in the literature review. We use the following notation:

n	number of observed historical demand values (strictly positive integer)
$f(d_j)$	population probability mass function for demand in period j (for $j = 1, ..., n$)
m	number of resampled demand values (strictly positive integer $(n \geq m \geq 2)$)
y	cumulative demand over m time periods
$F_m(y)$	population CDF of the cumulative demand over m time periods, evaluated at y
$\widehat{F}_m^{\mathrm{R}}(y)$	estimated CDF, over m time periods, by resampling with replacement (R), evaluated at y
$\widehat{F}_m^{\mathrm{NR}}(y)$	estimated CDF, over m time periods, by resampling 'without replacement (also known as 'no replacement', NR), evaluated at y.

Usually, the n historical observations will be for the most recent n time periods. However, this need not be the case. If the demand is genuinely i.i.d., then it will not matter if there are some missing historical values, and the n observed values stretch back further in time than n periods ago. Similarly, the cumulative demand, y, will usually relate to m successive time periods over the forthcoming lead time, as this is the most common inventory application. However, the cumulative demand may relate to a length of time that does not equate to the lead time, or to the total demand over time periods that are not successive, providing the i.i.d. assumption still holds.

2.2 Resampling without Replacement: Lack of CDF Bias

If m demand values are resampled with no replacement, then m distinct time indices are sampled from $1, ..., n$, say $i_1, ..., i_m$, assuming that there are sufficient historical observations available $(n \geq m)$. For convenience, the sampled indices are relabelled as $1, ..., m$, with the demand values associated with these indices being $d_1, ..., d_m$.

As all of the resampled observations are mutually independent, the expected value of the resampled CDF estimate, with no replacement, is given by:

$$\mathbb{E}(\widehat{F}_m^{\mathrm{NR}}(y)) = \sum_{d_1=0}^{y} \sum_{d_2=0}^{y-d_1} \dots \sum_{d_m=0}^{y-d_1-\dots-d_{m-1}} \prod_{j=1}^{m} f(d_j) = F_m(y) \tag{1}$$

It is evident that resampling with no replacement gives an unbiased estimate of the population cumulative distribution function, $F_m(y)$.

2.3 Resampling with Replacement: Existence and Quantification of CDF Bias

If aggregation is over only one period, then there is no need to distinguish between 'replacement' and 'no replacement'. In this case, resampling is an unbiased estimator of the cumulative distribution function.

When the length of aggregation extends to two periods or more, then resampling estimates with replacement are no longer unbiased. The bias, $\mathbb{E}(\widehat{F}_m^{\mathrm{R}}(y)) - F_m(y)$, can be found from Proposition 1 (in which $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x).

Proposition 1. *The expected value of the CDF estimate of the cumulative demand over m time periods, evaluated at y, for discrete i.i.d. demand with probability mass function f, using the method of resampling with replacement from n historical time periods, with demands $d_1, ..., d_n$, is given by:*

$$\mathbb{E}(\widehat{F}_m^{\mathrm{R}}(y)) = \frac{1}{n^m} \sum_{k=1}^{\min(m,n)} \sum_{\lambda \in \mathscr{P}(m,k)} N(n,m,k,\lambda) \left(\sum_{d_1=0}^{U_1} \dots \sum_{d_k=0}^{U_k} \prod_{j=1}^{k} f(d_j) \right) \tag{2}$$

where:

$$N(n,m,k,\lambda) = \frac{n!m!}{(n-k)!\lambda_1!...\lambda_k!r_1!...r_{|\lambda|}!}$$

$$U_1 = \lfloor y/\lambda_1 \rfloor, ..., U_k = \lfloor (y - \lambda_1 d_1 - ... - \lambda_{k-1} d_{k-1})/\lambda_k \rfloor$$

and $\mathscr{P}(m,k)$ is the set of integer partitions $\lambda = (\lambda_1, ..., \lambda_k)$ of m into k parts, where $\lambda_1 \geq ... \geq \lambda_k$ and there are $|\lambda|$ distinct values, $\lambda'_1 > ... > \lambda'_{|\lambda|}$ repeated $r_1, ..., r_{|\lambda|}$ times.

The proof of Proposition 1 is given in Appendix A. In this proposition, the integer partition $\lambda_1 + ... + \lambda_k = m$ shows how the total number of resampled time indices (m) splits into repeated selections. For example, for $m = 5$, the partition $\lambda = (2, 2, 1)$ indicates that one previous time index is resampled twice ($\lambda_1 = 2$), a second time index is resampled twice ($\lambda_2 = 2$), and a third time index is resampled once ($\lambda_3 = 1$). This gives two distinct values, $\lambda'_1 = 2$ and $\lambda'_2 = 1$, with the first distinct value repeated twice ($r_1 = 2$).

2.4 Resampling with Replacement: Asymptotic Result

Proposition 1, based on a finite length of demand history, leads to the following corollary:

Corollary For a fixed value of m, $\mathbb{E}(\widehat{F}^{\mathrm{R}}_m(y)) \to F_m(y)$ as $n \to \infty$ for all integer values of m greater than or equal to two.

In Equation (2), the first upper limit of summation is m, as consideration is being given to the limit as $n \to \infty$, for fixed m. The terms from $k = 1$ to $k = m-1$ and the term for $k = m$ are analyzed separately. For the former terms, $k \leq m-1$ and the expression $n!/(n-k)!$ has the highest power of n being no greater than $m-1$. After division by n^m, the limit of each of these terms (as $n \to \infty$) is zero ($k = 1, ..., m-1$). The latter term, $k = m$, corresponds to the partition $m = 1+1+...+1$ (and $r_1 = m$), and $U_1 = y, U_2 = y-d_1, ..., U_m = y-d_1-...-d_{m-1}$. For this term,

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[\widehat{F}^{\mathrm{R}}_m(y)] &= \lim_{n\to\infty} \frac{1}{n^m} \prod_{i=1}^{m}(n-i+1) \sum_{d_1=0}^{y} \sum_{d_2=0}^{y-d_1} ... \sum_{Y_m=0}^{y-d_1-...-d_{m-1}} \prod_{j=1}^{m} f(d_j) \\ &= \sum_{d_1=0}^{y} \sum_{d_2=0}^{y-d_1} ... \sum_{d_m=0}^{y-d_1-...-d_{m-1}} \prod_{j=1}^{m} f(d_j) = F_m(y)\end{aligned} \tag{3}$$

Hence, the estimator is aymptotically unbiased.

The implication of this corollary is that the bias of resampling with replacement becomes negligible for long demand histories. However, the assumption of an identical distribution of demand over time can become unsustainable as the demand history lengthens.

2.5 Magnitude of the bias by resampling with replacement

Equation (2) reduces to a much simpler form for an aggregation of demand over two periods ($m = 2$):

$$\mathbb{E}(\widehat{F}^{\mathrm{R}}_2(y)) = \frac{1}{n} F_1\left(\lfloor y/2 \rfloor\right) + \left(1 - \frac{1}{n}\right) F_2(y) \tag{4}$$

This equation has a natural interpretation. For $m = 2$, there are two possible ways in which resampling may occur: i) the same historical time index is sampled twice; ii) different historical time indices are sampled. If there are n historical demand observations, then the probabilities of these occurrences are $1/n$ and $1 - 1/n$, respectively, for random selections. In the latter case, there is no bias, as it is equivalent to resampling 'without replacement'. In the former case, the expected value of the CDF is found by identifying the highest value that can be chosen twice without the total exceeding the given value of y. This is clearly $\lfloor y/2 \rfloor$ for discrete demand and so the first term is specified accordingly.

The bias of the resampling estimate of the CDF (with replacement), for $m = 2$ and for discrete demand, may be written as follows:

$$\mathbb{E}(\widehat{F}^{\mathrm{R}}_2(y)) - F_2(y) = \frac{1}{n} \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \frac{1}{n} \sum_{i=0}^{y} \sum_{j=0}^{y-i} f(i) f(j) \tag{5}$$

The sign and magnitude of the bias depend on the form of the probability mass function and the cumulative demand over two periods (y). Suppose that demand over one period can take only three values: zero, one and two, and that it can take any probability mass function $f(0)$, $f(1)$ and $f(2) = 1 - f(0) - f(1)$. Then, the conditions for positive bias, using resampling with replacement over two periods, vary according to the value of y. For $y = 1$, the condition is that $f(0) < 1 - 2f(1)$ (provided $f(0) > 0$). For $y = 2$, the condition is that $f(0) < f(1)$ (provided $f(2) > 0$). Finally, for $y = 3$. the bias cannot be positive. The maximum and minimum bias values, and the conditions under which they are attained, are given in Proposition 2.

Proposition 2. *The resampled estimate (with replacement) of the cumulative distribution function (CDF), evaluated at y, for discrete i.i.d. demand with probabilty mass function f, when resampling two periods from n historical time periods, has a maximum bias of $1/4n$, for $y \geq 0$; and has a minimum bias of $-1/4n$, for $y \geq 1$, and zero for $y = 0$. The maximum bias of the CDF estimate is attained under the following conditions:*

$$\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) = \frac{1}{2}, \sum_{i=y+1}^{\infty} f(i) = 1 - \sum_{i=0}^{y} f(i) \tag{6}$$

and, if $y \geq 1$, for $i = \lfloor y/2 \rfloor + 1, ..., y$:

$$f(i) \sum_{j=0}^{y-i} f(j) = 0$$

The minimum bias of the CDF estimate, evaluated at y, is attained, for $y = 0$, when $f(0) = 0$ and $f(0) = 1$ with $\sum_{i=1}^{\infty} f(i) = 1 - f(0)$ in both cases. For $y \geq 1$, the minimum is attained under the following conditions:

$$\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) = \frac{1}{2} \sum_{i=\lfloor y/2 \rfloor+1}^{y} f(i) = \frac{1}{2} \tag{7}$$

and, for $i = 0, ..., \lfloor y/2 \rfloor$:

$$f(i) \sum_{j=0}^{y-i} f(j) = f(i)$$

and, for $i = \lfloor y/2 \rfloor + 1, ..., y$:

$$f(i) \sum_{j=0}^{y-i} f(j) = \frac{1}{2} f(i)$$

The proof of Proposition 2 is given in Appendix B. This result gives a formula for a precise bound of $1/4n$ on the absolute value of the bias of the resampling estimate (with replacement) for $m = 2$. Consequently, for longer demand histories, the bias will be modest. However, for shorter histories, often observed in practice, the bias becomes of greater concern.

Derivation of closed-form expressions for the maximum and minimum biases for aggregation levels of three periods or more ($m \geq 3$) becomes intractable. Numerical experimentation shows that the maximum and minimum biases may be of greater absolute value than those for $m = 2$. As an illustration, for an aggregation level of $m = 4$, and for $y = 3$, and for short histories of 10 and 20 observations, the maximum biases are 6.16% and

3.12%, respectively. This example shows that the bias in the estimation of the CDF may, in some cases, become of sufficient magnitude to be of practical significance, especially for shorter demand histories.

3 Variance of CDF Estimates by Resampling (With and Without Replacement)

In this section, exact analytical expressions are given for the variance of CDF estimates generated by resampling with and without replacement, for i.i.d. demand. The complexity of these expressions, particularly for 'with replacement', makes it very difficult to obtain simple conditions under which one method outperforms the other in terms of mean square error. However, in the final sub-section, examples are given to show that resampling with 'replacement' does not always dominate resampling with 'no replacement'.

3.1 Variance of CDF resampling estimate without replacement

Proposition 3. *The variance of the CDF estimate, $\widehat{F}_m^{\mathrm{NR}}(y)$, evaluated at y, for discrete i.i.d. demand, when resampling m periods with no replacement from n historical time periods ($n \geq m$), is given by the following formula:*

$$\mathrm{Var}(\widehat{F}_m^{\mathrm{NR}}(y)) = \sum_{k=0}^{m} \frac{\binom{m}{k}\binom{n-m}{m-k}}{\binom{n}{m}} \Theta_{k,m}(y) - F_m(y)^2 \tag{8}$$

where:

$$\Theta_{k,m}(y) = \sum_{d_1=0}^{y} \cdots \sum_{d_m=0}^{y-d_1-\ldots-d_{m-1}} \sum_{d_{m+1}=0}^{y-d_{m-k+1}-\ldots-d_m} \cdots \sum_{d_{2m-k}=0}^{y-d_{m-k+1}-\ldots-d_{2m-k-1}} \prod_{j=1}^{2m-k} f(d_j)$$

The proof of Proposition 3 is given in Appendix C.
Resampling 'without replacement' produces an unbiased estimate of the CDF. Therefore, the result given in Proposition 3 is also an expression for the mean square error.

In this proposition, $\Theta_{k,m}(y)$ ($k = 0, 1, ..., m$) represents the chance that two sets of resamplings ('without replacement') of m previous periods, which have an intersection of k periods, both have total demands not exceeding y. We note that, for all values of m, $\Theta_{0,m}(y) = F_m(y)^2$ (providing $n \geq 2m$), and for all values of m, $\Theta_{m,m}(y) = F_m(y)$. For the special case of $m = 2$ (and $n \geq 4$):

$$\mathrm{Var}(\widehat{F}_2^{\mathrm{NR}}(y)) = \frac{(n-2)(n-3)F_2(y)^2 + 4(n-2)\Theta_{1,2}(y) + 2F_2(y)}{n(n-1)} - F_2(y)^2 \tag{9}$$

where $\Theta_{1,2}(y) = \sum_{d_1=0}^{y} \sum_{d_2=0}^{y-d_1} \sum_{d_3=0}^{y-d_2} f(d_1)f(d_2)f(d_3)$ and $F_2(y) = \sum_{d_1=0}^{y} \sum_{d_2=0}^{y-d_1} f(d_1)f(d_2)$.

In the special case of $m = 3$ (and $n \geq 5$), the variance is given by:

$$\mathrm{Var}(\widehat{F}_3^{\mathrm{NR}}(y)) = \frac{(n-3)(n-4)(n-5)F_3(y)^2 + 9(n-3)(n-4)\Theta_{1,3}(y) + 18(n-3)\Theta_{2,3} + 6F_3(y)}{n(n-1)(n-2)} - F_3(y)^2 \tag{10}$$

where $\Theta_{1,3}(y)=\sum_{d_1=0}^{y}\sum_{d_2=0}^{y-d_1}\sum_{d_3=0}^{y-d_1-d_2}\sum_{d_4=0}^{y-d_3}\sum_{d_5=0}^{y-d_3-d_4} f(d_1)f(d_2)f(d_3)f(d_4)f(d_5)$

and $\Theta_{2,3}(y)=\sum_{d_1=0}^{y}\sum_{d_2=0}^{y-d_1}\sum_{d_3=0}^{y-d_1-d_2}\sum_{d_4=0}^{y-d_2-d_3} f(d_1)f(d_2)f(d_3)f(d_4)$

Similar expressions can be derived for $m \geq 4$ but the expressions become very lengthy.

3.2 Variance of CDF resampling estimate with replacement

Proposition 4. *The variance of the CDF estimate of the cumulative demand over m time periods, $\widehat{F}_m^{\mathrm{R}}(y)$ evaluated at y, for discrete i.i.d. demand, using the method of resampling with replacement from n historical time indices with demands $d_1, ..., d_n$, is given by:*

$$\mathrm{Var}(\widehat{F}_m^{\mathrm{R}}(y)) = \frac{1}{n^{2m}} \sum_{k=1}^{\min(2m,n)} \Bigg(\underset{|A|+|B-A|=k}{\sum_{\lambda^A}\sum_{\lambda^B}\sum_{\lambda^{B-A}}} N(n,m,k,\lambda^A,\lambda^B,\lambda^{B-A})\mathbb{E}(\mathbb{1}_{A^*,B^*}) \Bigg) - \mathbb{E}(\widehat{F}_m^{\mathrm{R}}(y))^2 \tag{11}$$

where $\mathbb{E}(\widehat{F}_m^{\mathrm{R}}(y))$ is given by Proposition 1, A^* and B^* are m-permutations (with repetition) from n time indices, A and B are the associated sets of distinct time indices, $\lambda^A = (\lambda_1^A, ..., \lambda_{|A|}^A)$ is a partition of m representing the multiplicities of the elements of A in A^* (and similarly for λ^B), $\lambda^{B-A} = (\lambda_1^{B-A}, ..., \lambda_{|B-A|}^{B-A})$ is a partition representing the multiplicities of the elements of $B-A$ in B^*, and $|\lambda^A|$ denotes the number of distinct elements in the partition λ^A, with $r_1^A, ..., r_{|\lambda^A|}^A$ indicating the number of repetitions of those distinct elements (and similarly for λ^{B-A}) and:

$$N(n,m,k,\lambda^A,\lambda^B,\lambda^{B-A}) = \frac{n!|A|!}{(n-k)!(|A|-|B-A|)!|B-A|!}\frac{1}{\Pi_{j=1}^{|\lambda^A|} r_j^A! \Pi_{j=1}^{|\lambda^{B-A}|} r_j^{B-A}!}\frac{(m!)^2}{\Pi_{j=1}^{|A|}\lambda_j^A! \Pi_{j=1}^{|B|}\lambda_j^B!}$$

$$\mathbb{E}(\mathbb{1}_{A^*,B^*}) = \sum_{d_1=0}^{U_1} \cdots \sum_{d_{\alpha+\beta+\gamma}=0}^{U_{\alpha+\beta+\gamma}} \prod_{j=1}^{\alpha+\beta+\gamma} f(d_j)$$

$$U_i = \begin{cases} \left\lfloor \left(y - \sum_{j=1}^{i-1} \lambda_j^A d_j\right) / \lambda_i^A \right\rfloor & i = 1, ..., \alpha \\ \left\lfloor \left(y - \sum_{j=\alpha+1}^{i-1} \lambda_j^B d_j\right) / \lambda_i^B \right\rfloor & i = \alpha+1, ..., \alpha+\beta \\ \min\left(\left\lfloor \left(y - \sum_{j=1}^{\alpha} \lambda_j^A d_j - \sum_{j=\alpha+\beta+1}^{i-1} \lambda_j^A d_j\right) / \lambda_i^A \right\rfloor, \left\lfloor \left(y - \sum_{j=\alpha+1}^{i-1} \lambda_j^B d_j\right) / \lambda_i^B \right\rfloor \right) & i = \alpha+\beta+1, ..., \alpha+\beta+\gamma \end{cases}$$

where, after renumbering, $A-B=\{d_1, ..., d_\alpha\}$, $B-A=\{d_{\alpha+1}, ..., d_{\alpha+\beta}\}$ *and* $A\cap B=\{d_{\alpha+\beta+1}, ..., d_{\alpha+\beta+\gamma}\}$

The proof of Proposition 4 is given in Appendix D.

The variance formula becomes lengthy as m increases but is of a shorter form for $m=2$. The components of the calculation of $n^4\mathbb{E}(\widehat{F}_2^{\mathrm{R}}(y)^2)$ for all feasible combinations of λ^A, λ^B and λ^{B-A} are shown in Table 1 (where $\Lambda_{1,2}(y) = \sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-i} f(d_i)f(d_j)$).

Table 1: Calculation of $n^4\mathbb{E}(\widehat{F}_2^{\mathrm{R}}(y)^2$

$k=\lvert A\cup B\rvert$	λ^A	λ^B	λ^{B-A}	$n^4\mathbb{E}(\widehat{F}_2^{\mathrm{R}}(y)^2)$
1	(2)	(2)		$nF_1(\lfloor y/2\rfloor)\ +$
2	(2)	(2)	(2)	$n(n-1)F_1(\lfloor y/2\rfloor)^2\ +$
2	(1,1)	(2)		$2n(n-1)\Lambda_{1,2}(y)\ +$
2	(2)	(1,1)	(1)	$2n(n-1)\Lambda_{1,2}(y)\ +$
2	(1,1)	(1,1)		$2n(n-1)F_2(y)\ +$
3	(1,1)	(2)	(2)	$n(n-1)(n-2)F_1(\lfloor y/2\rfloor)F_2(y)\ +$
3	(2)	(1,1)	(1,1)	$n(n-1)(n-2)F_1(\lfloor y/2\rfloor)F_2(y)\ +$
3	(1,1)	(1,1)	(1)	$4n(n-1)(n-2)\Theta_{1,2}(y)\ +$
4	(1,1)	(1,1)	(1,1)	$n(n-1)(n-2)(n-3)F_2(y)^2$

In Table 1, λ^{B-A} does not vary for any given combination of λ^A and λ^B, but may vary when $m \geq 3$. For example, if $\lambda^A = (3)$ and $\lambda^B = (2,1)$, then λ^{B-A} may be (2) or (1).

Using the results in Table 1 and expressing the probabilistic terms explicitly yields the following variance formula for $m = 2$:

$$\begin{aligned}\mathrm{Var}(\widehat{F}_2^R(y)) = \frac{1}{n^4}\Bigg[& n(n-1)\left(\sum_{i=0}^{\lfloor y/2\rfloor} f(i)\right)^2 + 2(n-2)\sum_{i=0}^{\lfloor y/2\rfloor} f(i)\sum_{i=0}^{y}\sum_{j=0}^{y-i} f(i)f(j) + (n-2)(n-3)\left(\sum_{i=0}^{y}\sum_{j=0}^{y-i} f(i)f(j)\right)^2 \\ &+ 4n(n-1)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-i} f(i)f(j) + 4n(n-1)(n-2)\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{l=0}^{y-j} f(i)f(j)f(l) \\ &+ n\sum_{i=0}^{\lfloor y/2\rfloor} f(i) + 2n(n-1)\sum_{i=0}^{y}\sum_{j=0}^{y-i} f(i)f(j)\Bigg] - \left[\frac{1}{n}\left(\sum_{i=0}^{\lfloor y/2\rfloor} f(i)\right) + (1-\frac{1}{n})\left(\sum_{i=0}^{y}\sum_{j=0}^{y-i} f(i)f(j)\right)\right]^2 \end{aligned} \quad (12)$$

The variance formula for $m = 3$ is tractable but lengthy; it is provided in Appendix E and used later in the numerical experiments.

3.3 With Replacement vs. Without Replacement

The results of Propositions 3 and 4 allow comparison of resampling with and without replacement. Numerical evaluation of the formulae in these propositions is sufficient to show that one method does not dominate the other, either in terms of the variance or mean square error of the CDF estimates. Consider the simplest special case, of m=2 and y=1, with intermittent i.i.d. demand, with probabilities of zero lead-time demand, $\mathbb{P}(0)$, and lead-time demand of one unit, $\mathbb{P}(1)$. $\mathbb{P}(0)$ and $\mathbb{P}(1)$ are varied between 0.1 and 0.9 with a step increase of 0.2. The two components of the mean square error (squared bias and variance) are evaluated in Tables 2 and 3, for demand histories of n=5 and n=100. Results for n=10 and n=20 are included in Tables 12 and 13 in Appendix F. Tables 2, 3, 12 and 13 show the differences in squared biases (Δ_1), and in variances (Δ_2), between resampling with replacement (R) and with no replacement (NR), both expressed as a percentage of the mean square error

(NR). The expressions for Δ_1 and Δ_2 are given by:

$\Delta_1 = 100((\text{Bias}^{\text{R}})^2 - (\text{Bias}^{\text{NR}})^2)/\text{MSE}^{\text{NR}}$ and $\Delta_2 = 100(\text{Var}^{\text{R}} - \text{Var}^{\text{NR}})/\text{MSE}^{\text{NR}}$.

The calculations in the relevant tables are based on the bias results in Sections 2.2 and 2.3, and the variance results in Sections 3.1 and 3.2. Positive results indicate a lower value for 'no replacement'; negative values indicate a lower value for 'replacement'.

Table 2: Percentage differences in squared biases (Δ_1) and variances (Δ_2) between 'replacement' and 'no replacement' ($m = 2$, $n = 5$)

		$\mathbb{P}(1)$				
$\mathbb{P}(0)$		0.1	0.3	0.5	0.7	0.9
0.1	Δ_1 (%)	3.65	0.24	0.01	0.24	0.55
	Δ_2 (%)	17.93	-2.97	-10.23	-14.40	-17.25
0.3	Δ_1 (%)	2.69	0.06	0.43	2.03	
	Δ_2(%)	-2.28	-10.81	-12.68	-12.37	
0.5	Δ_1 (%)	1.25	0.12	4.44		
	Δ_2 (%)	-11.09	-13.57	-4.00		
0.7	Δ_1 (%)	0.24	9.03			
	Δ_2 (%)	-15.71	13.89			
0.9	Δ_1 (%)	21.18				
	Δ_2 (%)	86.35				

Table 3: Percentage differences in squared biases (Δ_1) and variances (Δ_2) between 'replacement' and 'no replacement' ($m = 2$, $n = 100$)

		$\mathbb{P}(1)$				
$\mathbb{P}(0)$		0.1	0.3	0.5	0.7	0.9
0.1	Δ_1 (%)	0.29	0.02	0.00	0.01	0.03
	Δ_2 (%)	2.07	0.31	-0.32	-0.66	-0.89
0.3	Δ_1 (%)	0.16	0.00	0.02	0.11	
	Δ_2(%)	0.16	-0.41	-0.56	-0.57	
0.5	Δ_1 (%)	0.07	0.01	0.25		
	Δ_2 (%)	-0.49	-0.65	-0.01		
0.7	Δ_1 (%)	0.01	0.58			
	Δ_2 (%)	-0.81	1.30			
0.9	Δ_1 (%)	2.15				
	Δ_2 (%)	7.72				

In accordance with our earlier results. Δ_1 is always positive, because 'no replacement' (NR) is always unbiased and 'replacement' (R) is always biased. The differences in variances (Δ_2) show that one method does not always dominate the other, and the same applies to the difference in mean square errors ($\Delta_1 + \Delta_2$). Indeed, for some combinations of probabilities, the reduction in mean square error, by using 'no replacement' instead of ‘replacement’, can be quite substantial. This is illustrated in Table 2 by the case of $m = 2$, $n = 5$, $\mathbb{P}(0) = 0.9$, and $\mathbb{P}(1) = 0.1$. In Table 3, for $m = 2$ and $n = 100$, the differences are much smaller, as was anticipated for longer demand histories. Experiments were also conducted for the cases of $m = 3$ and $m = 4$ and similar insights were gained.

For further illustration, we compare the percentage differences in squared biases and variances between replacement and no replacement when the demand follows a compound Poisson distribution. As noted earlier, resampling methods may be used for standard distributions and the compound Poisson distribution is flexible in modelling fast and slow moving demands, with evidence of goodness-of-fit to real demand data (Lengu et al. (2013), Prak et al. (2021)). Table 4 shows the results for Δ_1 and Δ_2 under a Poisson-geometric distribution where the Poisson demand arrival rate $\lambda = 0.1, 0.4, 0.7, 1, 2, 3$ and the theta parameter of the geometric demand size distribution takes the values of $\theta = 0.1, 0.4, 0.7$.

Table 4: Percentage differences in squared biases and and variances between ’replacement’ and ’no replacement’ under Poisson-geometric distribution (m= 2)

		θ $(n = 5)$			θ $(n = 100)$		
λ		0.1	0.4	0.7	0.1	0.4	0.7
0.1	Δ_1 (%)	0,37	0,04	0,14	0,02	0,00	0,01
	Δ_2 (%)	-17,55	-17,68	-15,94	-0,91	-0,95	-0,90
0.4	Δ_1 (%)	1,59	0,28	0,05	0,08	0,01	0,00
	Δ_2 (%)	-12,91	-15,34	-15,07	-0,61	-0,78	-0,78
0.7	Δ_1 (%)	2,99	0,66	0,00	0,17	0,04	0,00
	Δ_2 (%)	-7,15	-12,33	-13,80	-0,23	-0,57	-0,66
1	Δ_1 (%)	4,58	1,16	0,11	0,27	0,07	0,01
	Δ_2 (%)	-0,05	-8,54	-11,69	0,28	-0,29	-0,49
2	Δ_1 (%)	10,59	3,35	1,22	0,86	0,24	0,08
	Δ_2 (%)	36,75	11,16	1,68	3,33	1,37	0,64
3	Δ_1 (%)	15,76	5,52	2,67	2,12	0,60	0,26
	Δ_2 (%)	107,67	46,46	26,75	10,59	5,20	3,34

In Table 4, the differences in variances (Δ_2) show that one method does not always dominate the other, and the same applies to mean square errors ($\Delta_1+\Delta_2$). The results also show that for some combinations of the distribution’s parameters, the reduction in mean square error, by using ‘no replacement’ instead of ‘replacement’, can be quite substantial. This is illustrated in Table 4 by the case of $n = 5$, $\lambda = 3$ and all selected values of θ). The effect reduces as the λ parameter decreases. The effect is also diminished when the length of demand history, n, increases, as shown in Table 4 for the case of $n = 100$.

We now analyse the impact of increasing the number of resampled periods m. We report in Table 5 the results of Δ_1 and Δ_2 under a Poisson-geometric distribution with the same numerical settings but with $m = 3$.

Table 5: Percentage differences in squared biases and and variances between 'replacement' and 'no replacement' under Poisson-geometric distribution (m= 3)

		θ (n= 5)			θ (n= 100)		
λ		0.1	0.4	0.7	0.1	0.4	0.7
0.1	Δ_1 (%)	1,28	0,22	0,17	0,07	0,01	0,01
	Δ_2 (%)	-29,88	-30,78	-28,75	-1,65	-1,80	-1,70
0.4	Δ_1 (%)	0,17	1,64	0,17	0,34	0,10	0,01
	Δ_2 (%)	-28,75	-25,17	-29,18	-0,46	-1,13	-1,44
0.7	Δ_1 (%)	9,25	3,84	1,59	0,68	0,26	0,10
	Δ_2 (%)	-0,07	-16,12	-23,83	1,10	-0,20	0,87
1	Δ_1 (%)	13,20	6,47	3,89	1,12	0,50	0,29
	Δ_2 (%)	21,65	-3,50	-14,40	3,16	1,10	0,09
2	Δ_1 (%)	24,75	15,37	12,82	3,66	2,09	1,66
	Δ_2 (%)	151,38	74,00	48,74	16,05	9,50	7,01
3	Δ_1 (%)	36,36	22,07	19,01	9,00	5,74	4,97
	Δ_2 (%)	480,55	259,02	199,53	48,81	30,96	25,15

The findings from Table 5, for the case of $m = 3$, are similar to those from Table 4. Again, they show that the reduction in mean square error can be substantial. It also shows that the effect can be greater than the case of $m = 2$.

4 Inventory Implications of CDF Estimates by Resampling (With and Without Replacement)

In this section, we analyze the implications of the resampling with and without replacement on inventory performance. To do so, we consider an order-up-to-level (OUTL) inventory control policy and we analyse the implications in calculating the OUTL using both approaches. We start by doing this for theoretically generated data. We generate demand series following a Poisson-geometric process with parameters λ and θ. We calculate the OUTL needed to achieve a target cycle service level, CSL using the theoretical distribution and the empirical distribution obtained with and without replacement. We consider 1000 series of length n and for each series the empirical distribution is built using 3000 (sampling) replications. The results (average OUTL across all series) are reported for n = 10, 15, 25, 50 and CSL =90%, 95%, 99%. Three lead-time values are considered to calculate the OUTL: L = 2, 4, 6. The results are reported in Tables 6, 7 and 8 where Boot^R and Boot^{NR} refer to bootstrapping with 'replacement' and 'no replacement', respectively.

Table 6: Theoretical and achieved order-up-to-level using bootstrapping, with and without replacement, under Poisson-geometric distribution (Leadtime = 2)

		$\lambda = 0.2$ and $\theta = 0.2$			$\lambda = 1.5$ and $\theta = 0.2$		
		CSL=90%	CSL=95%	CSL=99%	CSL=90%	CSL=95%	CSL=99%
	Theoretical	7	11	20	31	37	51
$n = 10$	Boot^R	7.290	7.845	12.175	28.934	33.204	41.851
	Boot^{NR}	7.267	7.647	9.660	28.366	32.099	38.126
$n = 15$	Boot^R	8.592	8.800	12.257	30.212	34.890	43.881
	Boot^{NR}	8.591	8.717	10.904	29.854	34.183	41.7927
$n = 25$	Boot^R	6.711	11.527	14.151	29.767	35.528	45.397
	Boot^{NR}	6.698	11.523	13.436	29.426	34.992	44.022
$n = 50$	Boot^R	7.436	11.195	16.406	29.754	36.296	47.709
	Boot^{NR}	7.390	11.051	16.134	29.652	35.997	47.179

Table 7: Theoretical and achieved order-up-to-level using bootstrapping, with and without replacement, under Poisson-geometric distribution (Leadtime = 4)

		$\lambda = 0.2$ and $\theta = 0.2$			$\lambda = 1.5$ and $\theta = 0.2$		
		CSL=90%	CSL=95%	CSL=99%	CSL=90%	CSL=95%	CSL=99%
	Theoretical	12	16	26	52	60	77
$n = 10$	Boot^R	9.669	12.808	16.022	48.057	54.418	64.606
	Boot^{NR}	9.483	9.660	10.257	45.140	48.611	54.272
$n = 15$	Boot^R	10.230	12.578	17.961	51.587	58.345	71.26
	Boot^{NR}	10.094	12.121	13.346	49.815	55.350	64.33
$n = 25$	Boot^R	12.216	14.274	19.275	49.681	56.553	69.757
	Boot^{NR}	12.142	13.959	18.113	48.732	54.835	65.504
$n = 50$	Boot^R	12.348	16.352	22.656	50.104	57.805	71.855
	Boot^{NR}	12.265	16.227	22.086	49.787	57.250	70.484

Table 8: Theoretical and achieved order-up-to-level using bootstrapping, with and without replacement, under Poisson-geometric distribution (Leadtime = 6)

		$\lambda = 0.2$ and $\theta = 0.2$			$\lambda = 1.5$ and $\theta = 0.2$		
		CSL=90%	CSL=95%	CSL=99%	CSL=90%	CSL=95%	CSL=99%
	Theoretical	16	21	32	72	82	101
$n = 10$	Boot^R	14.992	16.587	21.975	68.828	76.085	89.861
	Boot^{NR}	10.275	10.449	10.489	59.252	62.361	66.322
$n = 15$	Boot^R	13.541	18.149	22.212	68.811	77.303	92.099
	Boot^{NR}	12.808	13.358	14.548	63.522	68.542	76.719
$n = 25$	Boot^R	15.210	18.777	26.243	68.384	76.74	92.970
	Boot^{NR}	14.645	17.118	20.501	65.867	72.435	83.869
$n = 50$	Boot^R	16.887	20.354	28.523	69.291	78.22	95.388
	Boot^{NR}	16.647	19.699	25.97	68.086	76.144	90.930

These tables show that there is no uniform outperformance of one method by another. For a target Cycle Service Level of 90%, and Poisson-geometric parameters of $\lambda = 0.2$ and $\theta = 0.2$, the 'without replacement' method is able, for $n = 50$, to achieve a lower mean OUTL than 'with replacement', but still exceeding the theoretical OUTL. In other settings, the 'without replacement' method has lower mean OUTLs than the theoretical values.

The results in Tables 6, 7 and 8 are all based on data conforming to a Poisson-geometric distribution. Of course, real-world data is not always so well behaved, and so we now empirically analyse the inventory performance of the two approaches. To do so, we use a dataset related to the monthly demand of 3000 spare parts from the automotive industry with a demand history composed of 2 years (i.e. 24 periods). The demand data descriptive statistics are reported in Table 9. where we report for the demand intervals, demand sizes and demand per period, the minimum, the first, second and third percentiles and the maximum.

Table 9: Descriptive statistics of the demand data

	Demand Intervals		Demand Sizes		Demand per period	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
Min	1.043	0.209	1.000	0.000	0.542	0.504
25%ile	1.095	0.301	2.050	1.137	1.458	1.31
Median	1.263	0.523	2.886	1.761	2.333	1.922
75%ile	1.412	0.733	5.000	3.357	4.167	3.502
Max	2.000	1.595	193.750	101.415	129.167	122.746

Table 9 demonstrates that the series are a mix of higher and lower volume SKUs and include intermittent demand items.

In Table 10, we report the average and standard deviation of the difference for each SKU between the (achieved) order-up-to-levels calculated using with and without replacement (statistics reported across 3000 SKUs). The results are reported for two demand history lengths: $n = 13$ and 24, three target CSL values: $CSL = 90\%$, 95%, 99% and three lead-time values: $L = 2$, 4, 6.

Table 10: Difference in empirically achieved order-up-to-level results, subtracting resampling 'without replacement' from 'with replacement'

		$n = 13$			$n = 24$		
		CSL=90%	CSL=95%	CSL=99%	CSL=90%	CSL=95%	CSL=99%
$L = 2$	Mean	0.174	0.399	0.572	0.098	0.151	0.571
	St. Dev	1.214	1.743	3.076	1.506	1.180	2.854
$L = 4$	Mean	0.886	1.433	3.844	0.350	0.562	1.552
	St. Dev	1.744	2.826	11.676	1.495	2.174	4.959
$L = 6$	Mean	3.000	5.087	8.140	1.353	2.228	4.832
	St. Dev	5.556	12.684	18.003	2.859	4.478	11.833

The mean differences are somewhat greater in Table 10 than was observed in Tables 6, 7 and 8. There are also some large standard deviatons, indicating the need to probe inventory performance more deeply.

We evaluate the performance of the WSS bootstrapping method where resampling with (WSS^R) and without (WSS^{NR}) replacement are used for the sampling of the nonzero demands. The probability of a non-zero demand occurrence is estimated using a Markov transition matrix, and this does not induce any further biases. However, no jittering is used in this method in order to avoid conflation with the effect of the bias resulting from jittering (Rego and Mesquita (2015), Boylan and Syntetos (2021)). The effect of jittering bias on inventories is an important issue and worthy of further research.

The order-up-to-level policy is used to control the inventory. We use the first 13 periods to initialise the calculations and the performance evaluation is evaluated over the remaining periods. We calculate the resulting holding volumes, backordering volumes and achieved CSL when both bootstrapping approaches are used. Then, the results (averages across 3000 SKUs) are reported for three target CSL values, namely: CSL = 90%, 95%, 99% and three lead-time values: $L = 2$, 4, 6. The empirical results are summarised in Table 11.

Table 11: Inventory performance of resampling with and without replacement

			Holding volumes	Backlog volumes	Achieved CSL (%)
$L = 2$	WSS^{R}	CSL=90%	8.756	0.449	90.9
		CSL=95%	10.793	0.298	94.1
		CSL=99%	14.769	0.160	96.9
	WSS^{NR}	CSL=90%	8.739	0.456	90.8
		CSL=95%	10.582	0.308	93.8
		CSL=99%	13.956	0.177	96.5
$L = 4$	WSS^{R}	CSL=90%	13.332	0.465	91.0
		CSL=95%	16.175	0.315	93.8
		CSL=99%	22.352	0.154	97.0
	WSS^{NR}	CSL=90%	12.670	0.504	90.2
		CSL=95%	14.845	0.368	92.9
		CSL=99%	18.432	0.226	95.6
$L = 6$	WSS^{R}	CSL=90%	18.905	0.429	91.9
		CSL=95%	22.796	0.292	94.2
		CSL=99%	29.428	0.156	96.9
	WSS^{NR}	CSL=90%	17.379	0.521	90.5
		CSL=95%	19.136	0.411	92.4
		CSL=99%	22.454	0.278	94.8

The results in Table 11 show that neither of the bootstrapping methods can reach the higher target CSLs. However, they can both achieve the target of 90%. The 'no replacement' approach can do so with lower average holding volumes.

Furthermore, in order to analyse the trade-off between the inventory holding volumes and service performance, we provide the efficiency curves of both bootstrapping methods. We do so for three values of lead-time and three target CSLs as shown in Figures 1-2. Note that Figure 1 shows the efficiency for the combined holding volumes and backordering volumes whereas Figure 2 shows the efficiency of the holding volumes and the acheived CSL.

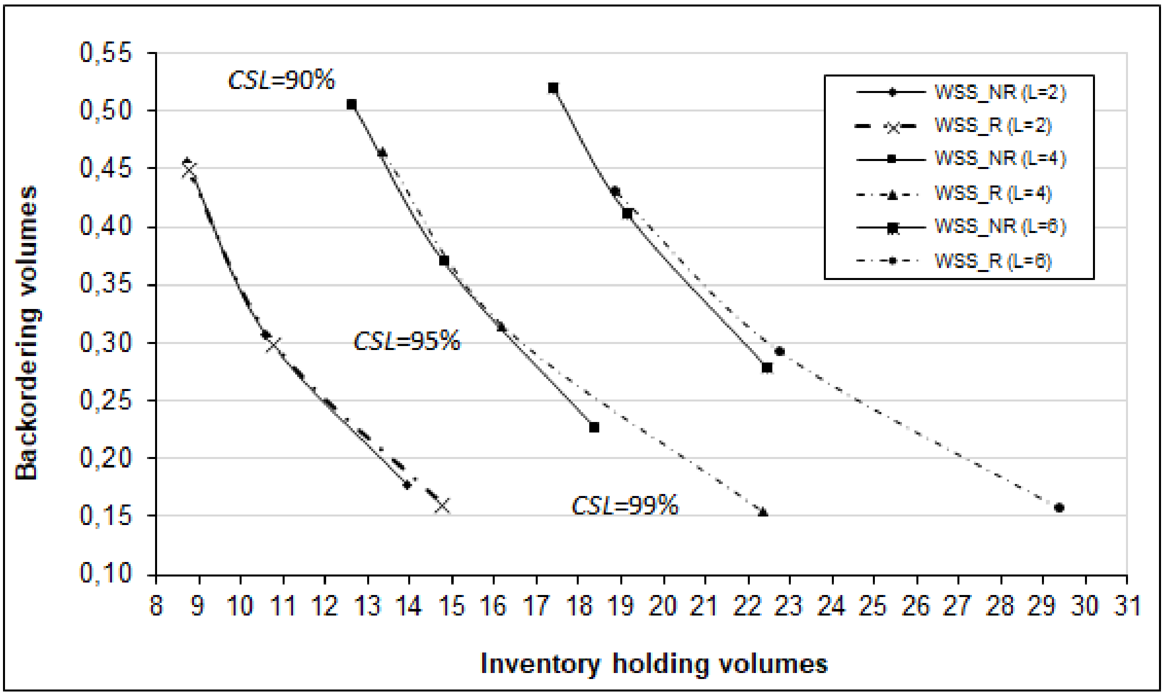

Figure 1: Efficiency curves (Holding volumes versus Backordering volumes) of WSS bootstrapping method with and without replacement (L = 2, 4, 6) and target CSL = 90%, 95%, 99%

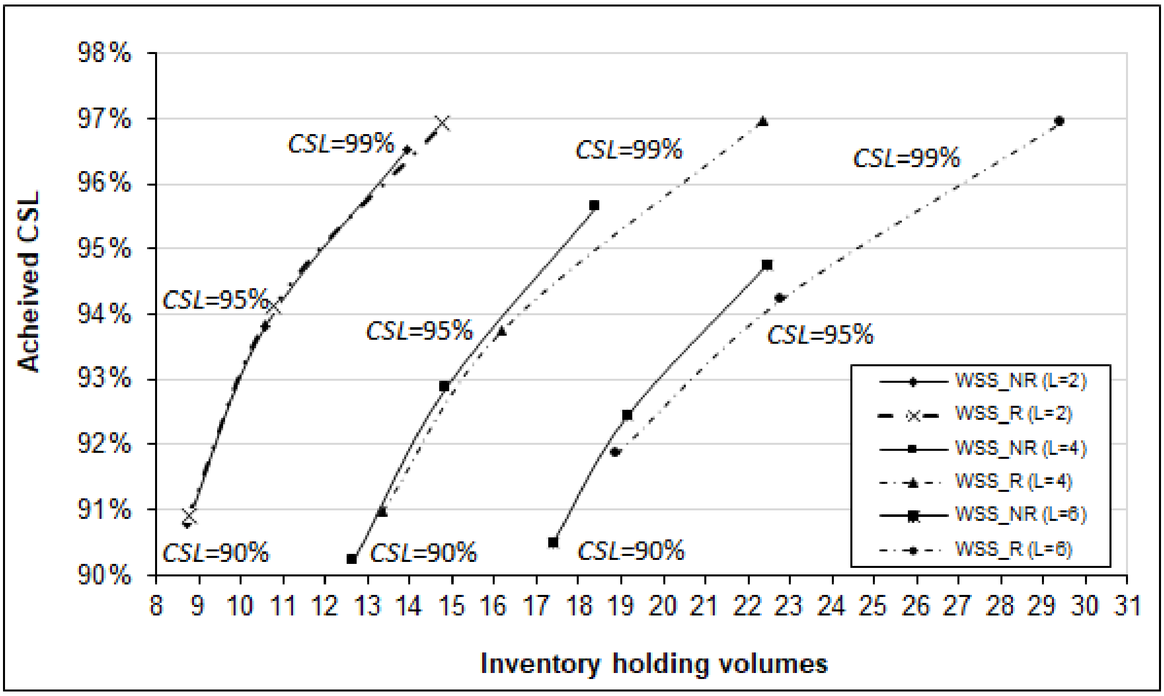

Figure 2: Efficiency curves (Holding volumes versus Acheived CSL) of WSS bootstrapping method with and without replacement (L = 2, 4, 6) and target CSL = 90%, 95%, 99%

Table 11 together with Figures 1 and 2, show that the WSS method (with Markov transition modelling but without jittering) can be improved by switching from ‘with replacement’ to ‘without replacement’ for a target Cycle Sevice Level of 90%. Reductions in inventory holding volumes of 0.2%, 5.0% and 8.1% are achieved by this switch, for lead times of 2, 4 and 6 periods, respectively. Consistent with the previous simulation results, the effect for short lead times is modest, but it becomes more pronounced for longer lead times. Also consistent with previous results is the finding that 'without replacement' falls further short than 'with replacement' for target Cycle Service Levels of 95% or 99%.

5 Conclusions

In previous research (e.g. Fricker and Goodhart (2000), Willemain et al. (2004), Zhou and Viswanathan (2011), Hasni et al. (2019)), it has been implicitly assumed that resampling with replacement should be used when resampling demands over single periods, in order to estimate the cumulative distribution function of lead-time demand. In this paper, we have evaluated the resampling with replacement approach by comparing it to resampling with no replacement. From a bias perspective, our conclusions are clear: resampling with replacement generates biased CDF estimates, whereas resampling with no replacement does not. We have shown that the severity of the bias (with replacement) depends on the value at which the CDF is to be estimated, the underlying probability distribution of demand, the lead time, and the length of demand history (n). This bias will flow through to inventory replenishment orders, which may be inflated if the CDF is underestimated or deflated if the CDF is overestimated, thus leading to excessive stocks or stock-outs. For a lead time of two periods, the maximum bias magnitide is $1/4n$, showing that the bias problem diminishes as the demand history lengthens. Indeed, we have proven that, for any length of lead time, resampling with replacement is asymptotically unbiased. Nevertheless, the bias issue is of some practical significance, given the short demand histories often encountered in industry.

We have given formulae for variances (and, hence, mean square errors) of the CDF estimates from reampling with replacement and with no replacement. It is worth noting that the bias and variance formulae derived in this paper can be modified to allow for continuous demand. However, product demand is almost invariably discrete and, therefore, resampling of discrete variables will be the norm in practice. Counter-intuitively, it has also been found that the variances of the CDF estimates from resampling with replacement are not always lower than with no replacement. The same is true for the mean square errors. The formulae for mean square errors of CDF estimates allow direct comparisons to be made between 'with replacement' and 'without replacement', especially for the simpler cases of short lead times of two or three periods. Even then, simulation of inventory performance is advisable to assess the stock implications of the two approaches. For longer lead times, the formulae for 'with replacement' become much lengthier, and simulation analyses may be conducted.

The empirical analysis conducted in this paper shows that, for a lower target Cycle Service Level (90%), a longer lead time (6 periods) and shorter demand histories (13 periods), a reduction in inventory holdings of over 8% is achievable by using 'without replacement', whilst still hitting the target CSL. However, 'with replacement' has been found to be preferable for the higher target CSLs of 95% and 99%.

Appendix A: Proof of Proposition 1 - Existence and quantification of CDF bias of resampling with replacement

As the resampling is with replacement, a time index may be sampled once, twice, or up to m times. Consequently, the number of distinct time indices, k, may take any integer value between 1 and m, provided that $n \geq m$; otherwise, if $n < m$, then k ranges from 1 to n. In general, k has a lower limit of 1 and an upper limit of $\min(m, n)$. Each value of k must be analyzed separately, to account for the varying number of repetitions of sampled time indices. The number of choices of the k distinct time indices, $\{i'_1, ..., i'_k\}$, from n historical time indices for which demand observations are available, is given by the standard binomial formula, $n!/(k!(n-k)!)$.

The restriction to k distinct time indices imposes the constraint, $\sum_{j=1}^{k} \lambda_j = \sum_{j=1}^{|\lambda|} r_j \lambda'_j = m$. Taking into account repetitions, there are $k!/(r_1!...r_{|\lambda|}!)$ permutations of the set $\{\lambda_1, ..., \lambda_k\}$, with the jth element in this set indicating the number of times the time index i'_j is resampled.

For a given allocation of the number of resamplings of $\{i'_1, ..., i'_k\}$, there are $m!/(\lambda_1!...\lambda_k!)$ arrangements of the time indices $\{i_1, ..., i_m\}$, taking repetitions into account. The product of the formulae for the number of choices of distinct time indices, the number of allocations of the amounts of resamplings, and the number of arrangements of the time indices, gives the result for $N(n, m, k, \lambda)$ shown in Proposition 1.

There are n^m ways of choosing m time indices from n indices (with replacement), and so the proportion of all possible choices of time indices that have a particular partition $\lambda = (\lambda_1, ..., \lambda_k)$ is given by $N(n, m, k, \lambda)/n^m$.

The expected value of the resampled (with replacement) estimate of the CDF, given an integer partition, $\lambda = (\lambda_1, ..., \lambda_k)$, is expressed as:

$$\mathbb{E}[\widehat{F}_m^{\mathrm{R}}(y)|\lambda] = \sum_{d_1=0}^{U_1} ... \sum_{d_k=0}^{U_k} \prod_{j=1}^{k} f(d_j) \tag{13}$$

where:

$$U_1 = \lfloor y/\lambda_1 \rfloor, U_2 = \lfloor (y - \lambda_1 d_1)/\lambda_2 \rfloor, ..., U_k = \lfloor (y - \lambda_1 d_1 - ... - \lambda_{k-1} d_{k-1})/\lambda_k \rfloor$$

Hence, weighting each of these expectations by the relevant proportions of integer partitions, and summing over all integer partitions with k parts, and summing over all possible values of k, establishes the result given in Proposition 1. $\square$

Appendix B: Proof of Proposition 2 - Resampling with replacement: maximization and minimization of bias (Aggregation of two periods)

In this appendix, the number of historical observations ($n \geq 2$) is kept fixed. The aim is to find the non-negative discrete distributions that maximize or minimize the bias (or, equivalently, n times the bias) for $m = 2$ and for any value of y. The bias is denoted by $B_{n,2}(y) = \mathbb{E}[\widehat{F}_2^{\mathrm{R}}(y)] - F_2(y)$.

Starting with the maximization problem, we analyse the cases $y = 0$ and $y \geq 1$ separately. For $y = 0$, we have: $nB_{n,2}(0) = f(0) - f(0)^2$, and it follows immediately that this is maximized at $f(0) = 1/2$ (and $\sum_{i=1}^{\infty} f(i) = 1/2$) and, for such a probability mass function, $B_{n,2}(0) = 1/4n$. For $y \geq 1$, the maximization problem is to choose $f(0), ..., f(y) \geq 0$ to maximize:

$$nB_{n,2}(y) = \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \sum_{i=0}^{y} \sum_{j=0}^{y-i} f(i)f(j) \tag{14}$$

subject to $\sum_{i=0}^{y} f(i) \leq 1$. The Lagrangian ($L$) is given by:

$$L = \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \sum_{i=0}^{y} \sum_{j=0}^{y-i} f(i)f(j) + \lambda \left(1 - \sum_{i=0}^{y} f(i)\right) \tag{15}$$

where $\lambda \geq 0$. This yields the following Karush-Kuhn-Tucker (KKT) conditions for $i = 0$ to $i = \lfloor y/2 \rfloor$:

$$f(i) \left(1 - 2\sum_{j=0}^{y-i} f(j) - \lambda\right) = 0 \tag{16}$$

and, for $i = \lfloor y/2 \rfloor + 1$ to $i = y$:

$$f(i) \left(-2\sum_{j=0}^{y-i} f(j) - \lambda\right) = 0$$

Therefore, for $i = \lfloor y/2 \rfloor + 1, ..., y$, either (i) $f(i) = 0$ or (ii) $\lambda = -2\sum_{j=0}^{y-i} f(j)$. If the second condition holds for a particular value of i, then $\lambda \geq 0$ implies that $\lambda = f(0) = f(1) = ... = f(y-i) = 0$. If the second condition does not hold, then $f(i) = 0$. Thus, for all i in the range $\lfloor y/2 \rfloor + 1 \leq i \leq y$:

$$f(i) \sum_{j=0}^{y-i} f(j) = 0 \tag{17}$$

Hence the objective function simplifies to:

$$\begin{aligned} nB_{n,2}(y) &= \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \sum_{i=0}^{\lfloor y/2 \rfloor} \sum_{j=0}^{y-i} f(i)f(j) \\ &= \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \sum_{i=0}^{\lfloor y/2 \rfloor} \sum_{j=0}^{\lfloor y/2 \rfloor} f(i)f(j) - \sum_{i=0}^{\lfloor y/2 \rfloor} \sum_{j=\lfloor y/2 \rfloor+1}^{y-i} f(i)f(j) \end{aligned} \tag{18}$$

From Equation 17, if $i \geq \lfloor y/2 \rfloor + 1$ and $i + j \leq y$, then $f(i)f(j) = 0$. Swapping the roles of i and j results in the final term of the last expression vanishing:

$$\begin{aligned} nB_{n,2}(y) &= \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \sum_{i=0}^{\lfloor y/2 \rfloor} \sum_{j=0}^{\lfloor y/2 \rfloor} f(i)f(j) \\ &= F_1\left(\lfloor y/2 \rfloor\right) - F_1\left(\lfloor y/2 \rfloor\right)^2 \end{aligned} \tag{19}$$

This is maximized at $F_1(\lfloor y/2 \rfloor) = 1/2$ at which:

$$B_{n,2}(y) = \frac{1}{4n} \tag{20}$$

Thus, $f(0), ..., f(\lfloor y/2 \rfloor)$ can be chosen in any way such that their sum is 1/2 and $f(\lfloor y/2 \rfloor + 1), ..., f(y)$ can be chosen in any way that satisfies Equation 17. Because f is a probability mass function, the final condition is that: $\sum_{i=y+1}^{\infty} f(i) = 1 - \sum_{i=0}^{y} f(i)$. For any distribution satisfying these conditions, the bias is maximised at $1/4n$, for all $y \geq 1$.

Now, we turn to the problem of finding the non-negative discrete distributions that minimize the bias (or, equivalently, n times the bias) for an aggregation level of $m = 2$ and for any value of y. Again, we analyze the cases $y = 0$ and $y \geq 1$ separately. For $y = 0$, we have: $nB_{n,2}(0) = f(0) - f(0)^2$, and it follows immediately that this is minimized at $f(0) = 0$ (with $\sum_{i=1}^{\infty} f(i) = 1$) and at $f(0) = 1$ (with $\sum_{i=1}^{\infty} f(i) = 0$). At both values, the bias takes a minimum value of $B_{n,2}(0) = 0$. For $y \geq 1$, the Lagrangian (L) for this minimization problem is given by:

$$L = -\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) + \sum_{i=0}^{y} \sum_{j=0}^{y-i} f(i)f(j) + \lambda \left(1 - \sum_{i=0}^{y} f(i)\right) \tag{21}$$

where $\lambda \geq 0$. This yields the following KKT conditions for $i = 0$ to $i = \lfloor y/2 \rfloor$:

$$f(i) \left(-1 + 2\sum_{j=0}^{y-i} f(j) - \lambda\right) = 0 \tag{22}$$

and for $i = \lfloor y/2 \rfloor + 1$ to $i = y$;

$$f(i) \left(2\sum_{j=0}^{y-i} f(j) - \lambda\right) = 0$$

If $\lambda = 0$, the second KKT condition reduces to Equation 17 and, as previously shown, the bias can then be calculated from Equation 19. This function is non-negative for any cumulative distribution function $F_1(\lfloor y/2 \rfloor)$. However, for $y \geq 1$, this cannot be the minimum value of $nB_{n,2}(y)$ because the bias can take negative values (for example, if $f(0) = f(y) = 1/2$). Hence, $\lambda > 0$ and the constraint in the original minimization problem is active, so that: $F_1(y) = \sum_{i=0}^{y} f(i) = 1$. The KKT conditions can be re-expressed as follows, summing from $i = 0$ to $i = \lfloor y/2 \rfloor$:

$$\sum_{i=0}^{\lfloor y/2 \rfloor} \sum_{j=0}^{y-i} f(i)f(j) = \frac{\lambda + 1}{2} \sum_{i=0}^{\lfloor y/2 \rfloor} f(i) \tag{23}$$

and summing from $i = \lfloor y/2 \rfloor + 1$ to $i = y$:

$$\sum_{i=\lfloor y/2 \rfloor+1}^{y} \sum_{j=0}^{y-i} f(i)f(j) = \frac{\lambda}{2} \sum_{i=\lfloor y/2 \rfloor+1}^{y} f(i) \tag{24}$$

Adding:

$$\sum_{i=0}^{y}\sum_{j=0}^{y-i} f(i)f(j) = \frac{\lambda}{2} + \frac{1}{2}\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) \tag{25}$$

This yields the following result:

$$nB_{n,2}(y) = \frac{1}{2}\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \frac{\lambda}{2} = \frac{1}{2}F_1\left(\lfloor y/2 \rfloor\right) - \frac{\lambda}{2} \tag{26}$$

Now, subtracting Equation 24 from Equation 23:

$$\sum_{i=0}^{\lfloor y/2 \rfloor}\sum_{j=0}^{\lfloor y/2 \rfloor} f(i)f(j) = \frac{\lambda+1}{2}\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) - \frac{\lambda}{2}\sum_{i=\lfloor y/2 \rfloor+1}^{y} f(i) \tag{27}$$

where, on the left-hand side, we have made use of the identity $\sum_{i=\lfloor y/2 \rfloor+1}^{y}\sum_{j=0}^{y-i} f(i)f(j) = \sum_{i=0}^{\lfloor y/2 \rfloor}\sum_{j=\lfloor y/2 \rfloor+1}^{y-i} f(i)f(j)$. Hence:

$$F_1\left(\lfloor y/2 \rfloor\right)^2 = \frac{\lambda+1}{2}F_1\left(\lfloor y/2 \rfloor\right) - \frac{\lambda}{2}\left(1 - F_1\left(\lfloor y/2 \rfloor\right)\right) \tag{28}$$

$$2F_1\left(\lfloor y/2 \rfloor\right)^2 - (2\lambda+1)F_1\left(\lfloor y/2 \rfloor\right) + \lambda = 0 \tag{29}$$

Solving this quadratic equation:

$$F_1\left(\lfloor y/2 \rfloor\right) = \frac{2\lambda + 1 \pm (2\lambda - 1)}{4} \tag{30}$$

This gives two solutions, namely $F_1\left(\lfloor y/2 \rfloor\right) = \lambda$ and $F_1\left(\lfloor y/2 \rfloor\right) = 1/2$.

Substituting into Equation 26, these solutions correspond to $nB_{n,2}(y) = 0$ and $nB_{n,2}(y) = 1/4 - \lambda/2$. It has already been noted that the minimum bias must be negative, which means that, for $y \geq 1$, the first solution cannot correspond to the minimum. We have shown previously that $F_1(y) = 1$ and since, for the second solution, $F_1(\lfloor y/2 \rfloor) = 1/2$, it follows that:

$$\sum_{i=0}^{\lfloor y/2 \rfloor} f(i) = \sum_{i=\lfloor y/2 \rfloor+1}^{y} f(i) = \frac{1}{2} \tag{31}$$

Thus, the bias is minimized if $f(0), ..., f(\lfloor y/2 \rfloor)$ and $f(\lfloor y/2 \rfloor + 1), ..., f(y)$ are chosen in any way such that both of their sums are $1/2$ and Equations 23 and 24 are satisfied. Because f is a probability mass function, $f(i) = 0$ for $i \geq y + 1$.

The pair $f(0) = f(y) = 1/2$ is a solution provided that the following constraints are satisfied:

$$\frac{1}{2} = f(0)^2 + f(0)f(y) = \frac{\lambda+1}{2}f(0) = \frac{\lambda+1}{4} \tag{32}$$

$$\frac{1}{4} = f(0)f(y) = \frac{\lambda}{2}f(y) = \frac{\lambda}{4} \tag{33}$$

Both of these constraints are satisfied at $\lambda = 1$. Thus one of the minimum solutions has been found, at which $nB_{n,2}(y) = 1/4 - 1/2 = -1/4$. This shows that the minimum value attained by the bias is given by:

$$B_{n,2}(y) = -\frac{1}{4n} \tag{34}$$

The constraint given by Equation 23, for $i = 0, ..., \lfloor y/2 \rfloor$ becomes, for $\lambda = 1$:

$$f(i) \sum_{j=0}^{y-i} f(j) = f(i) \tag{35}$$

The constraint given by Equation 24, for $i = \lfloor y/2 \rfloor + 1, ..., y$ becomes:

$$f(i) \sum_{j=0}^{y-i} f(j) = \frac{1}{2} f(i) \tag{36}$$

Appendix C: Proof of Proposition 3 - Variance expression of CDF resampling estimate without replacement

By definition,

$$\widehat{F}_m^{\mathrm{NR}}(y)^2 = \left(\frac{1}{\binom{n}{m}} \sum_{I^* \in \mathscr{I}} \mathbb{1}\left[\sum_{i \in I} d_i \le y \right] \right)^2 \tag{37}$$

where I^* represents a specific permutation of m indices that have been resampled (and $\mathscr{I}$ is the set of all such m-permutations from n indices that can be resampled without replacement) and I is the associated set of elements of the m-permutation. Then:

$$\widehat{F}_m^{\mathrm{NR}}(y)^2 = \frac{1}{\binom{n}{m}^2} \sum_{I^* \in \mathscr{I}} \sum_{J^* \in \mathscr{I}} \mathbb{1}\left[\sum_{i \in I} d_i \le y \right] \mathbb{1}\left[\sum_{j \in J} d_j \le y \right] \tag{38}$$

$$\mathbb{E}(\widehat{F}_m^{\mathrm{NR}}(y)^2) = \frac{1}{\binom{n}{m}^2} \sum_{k=0}^{m} \sum_{\substack{I^* \in \mathscr{I} \\ |I \cap J| = k}} \sum_{J^* \in \mathscr{I}} \mathbb{E}\left(\mathbb{1}\left[\sum_{i \in I} d_i \le y \right] \mathbb{1}\left[\sum_{j \in J} d_j \le y \right] \right) \tag{39}$$

Recalling the definition of $\Theta_{k,m}(y)$ for $k = 0, 1, ..., m$, it follows that :

$$\mathbb{E}(\widehat{F}_m^{\mathrm{NR}}(y)^2) = \frac{1}{\binom{n}{m}^2} \sum_{k=0}^{m} \sum_{\substack{I^* \in \mathscr{I} \\ |I \cap J| = k}} \sum_{J^* \in \mathscr{I}} \Theta_{k,m}(y) \tag{40}$$

Now, it is required to find the number of ways of choosing m elements for one set (with no replacement) and m elements for a second set (with no replacement) in such a way that k of them are in common. The total number of ways of choosing any m elements from n, for the first set, is $\binom{n}{m}$. The number of ways of choosing the elements of the second set so that k of them intersect, and $m-k$ do not, is $\binom{m}{k}\binom{n-m}{m-k}$. Hence, the unconditional expectation is given by:

$$\mathbb{E}(\widehat{F}_m^{\mathrm{NR}}(y)^2) = \sum_{k=0}^{m} \frac{\binom{m}{k}\binom{n-m}{m-k}}{\binom{n}{m}} \Theta_{k,m}(y) \tag{41}$$

By applying the standard identity $\mathrm{Var}(\widehat{F}_m^{\mathrm{NR}}(y)) = \mathbb{E}((\widehat{F}_m^{\mathrm{NR}}(y))^2) - (\mathbb{E}(\widehat{F}_m^{\mathrm{NR}}(y)))^2$, and recalling that $\mathbb{E}(\widehat{F}_m^{\mathrm{NR}}(y)) = F_m(y)$, Proposition 3 is established. □

Appendix D: Proof of Proposition 4 - Variance expression of CDF resampling estimate with replacement

By definition,

$$\widehat{F}_m^{\mathrm{R}}(y)^2 = \left(\frac{1}{n^m} \sum_{A^* \in \mathscr{A}} \mathbb{1}\left[\sum_{a \in A} \lambda_a^A d_a \leq y\right]\right)^2 = \frac{1}{n^{2m}} \sum_{A^* \in \mathscr{A}} \sum_{B^* \in \mathscr{A}} \mathbb{1}_{A^*,B^*} \tag{42}$$

where $\mathbb{1}_{A^*,B^*} = \mathbb{1}\left[\sum_{a\in A} \lambda_a^A d_a \leq y\right] \mathbb{1}\left[\sum_{b \in B} \lambda_b^B d_b \leq y\right]$ and $\mathscr{A}$ is the set of all m-permutations (with repetition) from n indices.

The expectation of $\mathbb{1}_{A^*,B^*}$ depends on the number of distinct elements selected over both sets A and B (namely $|A \cup B| = |A| + |B - A| = k$), which ranges from 1 to $\min(2m, n)$, and so:

$$\mathbb{E}(\widehat{F}_m^{\mathrm{R}}(y)^2) = \frac{1}{n^{2m}} \sum_{k=1}^{\min(2m,n)} \sum_{\substack{A^* \in \mathscr{A} \\ |A|+|B-A|=k}} \sum_{B^* \in \mathscr{A}} \mathbb{E}(\mathbb{1}_{A^*,B^*}) \tag{43}$$

Renumbering indices, we let $A - B = \{d_1, ..., d_\alpha\}$, $B - A = \{d_{\alpha+1}, ..., d_{\alpha+\beta}\}$ and $A \cap B = \{d_{\alpha+\beta+1}, ..., d_{\alpha+\beta+\gamma}\}$. Then the expectation of $\mathbb{1}_{A^*,B^*}$ may be written:

$$\mathbb{E}(\mathbb{1}_{A^*,B^*}) = \sum_{d_1=0}^{U_1} ... \sum_{d_\alpha=0}^{U_\alpha} \sum_{d_{\alpha+1}=0}^{U_{\alpha+1}} ... \sum_{d_{\alpha+\beta}=0}^{U_{\alpha+\beta}} \sum_{d_{\alpha+\beta+1}=0}^{U_{\alpha+\beta+1}} ... \sum_{d_{\alpha+\beta+\gamma}=0}^{U_{\alpha+\beta+\gamma}} \prod_{j=1}^{\alpha+\beta+\gamma} f(d_j) \tag{44}$$

where

$$U_1 = \lfloor y/\lambda_1^A \rfloor, ..., U_\alpha = \left\lfloor \left(y - \sum_{j=1}^{\alpha-1} \lambda_i^A d_i\right)/\lambda_\alpha^A \right\rfloor$$

$$U_{\alpha+1} = \lfloor y/\lambda_{\alpha+1}^B \rfloor, ..., U_{\alpha+\beta} = \left\lfloor \left(y - \sum_{i=\alpha+1}^{\alpha+\beta-1} \lambda_i^B d_i\right)/\lambda_{\alpha+\beta}^B \right\rfloor$$

$$U_{\alpha+\beta+1} = \min\left(\left\lfloor \left(y - \sum_{i=1}^{\alpha} \lambda_i^A d_i\right)/\lambda_{\alpha+\beta+1}^A \right\rfloor, \left\lfloor \left(y - \sum_{i=\alpha+1}^{\alpha+\beta} \lambda_i^B d_i\right)/\lambda_{\alpha+\beta+1}^B \right\rfloor\right), ...,$$

$$U_{\alpha+\beta+\gamma} = \min\left(\left\lfloor \left(y - \sum_{i=1}^{\alpha} \lambda_i^A d_i - \sum_{i=\alpha+\beta+1}^{\alpha+\beta+\gamma-1} \lambda_i^A d_i\right)/\lambda_{\alpha+\beta+\gamma}^A \right\rfloor, \left\lfloor \left(y - \sum_{i=\alpha+1}^{\alpha+\beta+\gamma-1} \lambda_i^B d_i\right)/\lambda_{\alpha+\beta+\gamma}^B \right\rfloor\right)$$

This shows that the expectation of $\mathbb{1}_{A^*,B^*}$ depends on λ^A and λ^B. The number of feasible permutations depends on the parameters n, m and k. It also depends on the partitions λ^A and λ^B, and on the partition λ^{B-A}, as will be discussed shortly. Hence, letting $N(n,m,k,\lambda^A,\lambda^B,\lambda^{B-A})$ represent the number of feasible permutations for given parameters and partitions:

$$\mathbb{E}(\widehat{F}^{\mathrm{R}}_m(y)^2) = \frac{1}{n^{2m}} \sum_{k=1}^{\min(2m,n)} \sum_{\substack{\lambda^A \\ |A\cup B|=k}} \sum_{\lambda^B} \sum_{\lambda^{B-A}} N(n,m,k,\lambda^A,\lambda^B,\lambda^{B-A}) \mathbb{E}(\mathbb{1}_{A^*,B^*}) \tag{45}$$

Now it is required to evaluate $N(n,m,k,\lambda^A,\lambda^B,\lambda^{B-A})$. Consideration needs to be given to three components of the calculation: i) allocation of time indices to the multiplicities of elements $\lambda^A_1,...,\lambda^A_{|A|}$; ii) selection from set A of those time indices to be members of set B (i.e. the elements of $A\cap B$); and iii) allocation of time indices to the mutiplicities of elements $\lambda^{B-A}_1,...,\lambda^{B-A}_{|B-A|}$.

For the first component, the number of permutations is given by the following expression, allowing for repetition of multiplicities: $n!/((n-|A|)!\Pi_{j=1}^{|\lambda^A|} r^A_j!)$; a similar expression applies for the third component, namely $(n-|A|)!/((n-|A|-|B-A|)!\Pi_{j=1}^{|\lambda^{B-A}|} r^{B-A}_j!)$ (noting that $|A|+|B-A| = |A\cup B|$). For the second component, the number of choices of $|A\cap B|$ intersecting elements from the set A is given by the standard binomial formula: $|A|!/(|A\cap B|!(n-|A\cap B|)!)$.

Finally, the number of arrangements of the time-indices in sets A and B must be quantified. taking into account the multiplicities for sets A and B. The formulae for the arrangements are: $m!/\Pi_{j=1}^{|A|}\lambda^A_j!$ and $m!/\Pi_{j=1}^{|B|}\lambda^B_j!$ respectively.

Multiplying all the relevant components yields the expression for $N(n,m,k,\lambda^A,\lambda^B,\lambda^{B-A})$ given in Proposition 4. Therefore, the proposition is proven. $\square$

Appendix E: Variance expression of resampling with replacement (case of $m = 3$)

The following expression is a special case of the general formula (11).

$$
\begin{aligned}
\mathrm{Var}(\widehat{F}_3^R(y)) = \frac{1}{n^6}\Bigg[& n(n-1)(\sum_{i=0}^{\lfloor y/3\rfloor} f(i))^2 + 6n(n-1)(n-2)\sum_{i=0}^{\lfloor y/3\rfloor} f(i)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i} f(i)f(j) \\
& +2n(n-1)(n-2)(n-3)(\sum_{i=0}^{\lfloor y/3\rfloor} f(i))(\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k)) + 9n(n-1)(n-2)(n-3)(\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i} f(i)f(j))^2 \\
& +6n(n-1)(n-2)(n-3)(n-4)(\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i} f(i)f(j))(\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k)) \\
& +n(n-1)(n-2)(n-3)(n-4)(n-5)(\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k))^2 + 6n(n-1)\sum_{i=0}^{\lfloor y/3\rfloor}\sum_{j=0}^{\lfloor (y-i)/2\rfloor} f(i)f(j) \\
& +6n(n-1)(n-2)\sum_{i=0}^{\lfloor y/3\rfloor}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k) + 9n(n-1)(n-2)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i}\sum_{k=0}^{\lfloor (y-i)/2\rfloor} f(i)f(j)f(k) \\
& +9n(n-1)(n-2)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{Min(y-2i,\lfloor y/2\rfloor)}\sum_{k=0}^{y-2j} f(i)f(j)f(k) + 9n(n-1)(n-2)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i}\sum_{k=0}^{\lfloor (y-j)/2\rfloor} f(i)f(j)f(k) \\
& +18n(n-1)(n-2)(n-3)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i}\sum_{k=0}^{y-i}\sum_{l=0}^{y-i-k} f(i)f(j)f(k)f(l) + 6n(n-1)\sum_{i=0}^{\lfloor y/3\rfloor}\sum_{j=0}^{y-2i} f(i)f(j) \\
& +18n(n-1)(n-2)(n-3)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i}\sum_{k=0}^{y-j}\sum_{l=0}^{y-j-k} f(i)f(j)f(k)f(l) + 9n(n-1)(n-2)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i}\sum_{k=0}^{y-2i} f(i)f(j)f(k) \\
& +9n(n-1)(n-2)(n-3)(n-4)\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j}\sum_{l=0}^{y-i}\sum_{t=0}^{y-i-l} +9n(n-1)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{Min(y-2i,\lfloor (y-2i)/2\rfloor)} f(i)f(j) \\
& +36n(n-1)(n-2)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k) + 18n(n-1)(n-2)(n-3)\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j}\sum_{l=0}^{y-i-j} f(i)f(j)f(k)f(l) \\
& +n\sum_{i=0}^{\lfloor y/3\rfloor} f(i) + 9n(n-1)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i} f(i)f(j) + 6n(n-1)(n-2)\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k)\Bigg] \\
& -\frac{1}{n^3}\Bigg[n\sum_{i=0}^{\lfloor y/3\rfloor} f(i) + 3n(n-1)\sum_{i=0}^{\lfloor y/2\rfloor}\sum_{j=0}^{y-2i} f(i)f(j) + n(n-1)(n-2)\sum_{i=0}^{y}\sum_{j=0}^{y-i}\sum_{k=0}^{y-i-j} f(i)f(j)f(k)\Bigg]^2 \qquad (46)
\end{aligned}
$$

Appendix F: Bias and variance reduction - Case of discrete distribution ($n = 10, 20$)

Table 12: Percentage differences in squared biases (Δ_1) and variances (Δ_2) between replacement and no replacement ($n = 10$)

		$\mathbb{P}(1)$				
$\mathbb{P}(0)$		0.1	0.3	0.5	0.7	0.9
0.1	Δ_1 (%)	2.33	0.14	0.01	0.12	0.28
	Δ_2 (%)	14.36	0.78	-4.11	-6.90	-8.75
0.3	Δ_1 (%)	1.49	0.03	0.23	1.05	
	Δ_2(%)	0.27	-4.74	-5.93	-5.94	
0.5	Δ_1 (%)	0.65	0.07	2.37		
	Δ_2 (%)	-5.19	-6.63	-1.00		
0.7	Δ_1 (%)	0.12	5.16			
	Δ_2 (%)	-7.98	10.02			
0.9	Δ_1 (%)	15.00				
	Δ_2 (%)	57.67				

Table 13: Percentage differences in squared biases (Δ_1) and variances (Δ_2) between replacement and no replacement ($n = 20$)

		$\mathbb{P}(1)$				
$\mathbb{P}(0)$		0.1	0.3	0.5	0.7	0.9
0.1	Δ_1 (%)	1.32	0.07	0.00	0.06	0.14
	Δ_2 (%)	8.83	1.02	-1.79	-3.37	-4.41
0.3	Δ_1 (%)	0.78	0.018	0.12	0.530	
	Δ_2(%)	0.50	-2.199	-2.86	-2.914	
0.5	Δ_1 (%)	0.33	0.033	1.22		
	Δ_2 (%)	-2.51	-3.277	-0.25		
0.7	Δ_1 (%)	0.06	2.75			
	Δ_2 (%)	-4.02	5.82			
0.9	Δ_1 (%)	9.10				
	Δ_2 (%)	33.71				

Appendix G: Bias and variance reduction - Case of Poisson-geometric distribution (n = 10, 20)

Table 14: Percentage differences in squared biases and and variances between replacement and no replacement under Poisson-geometric distribution

		θ $(n = 10)$			θ $(n = 20)$		
λ		0.1	0.4	0.7	0.1	0.4	0.7
0.1	Δ_1 (%)	0,18	0,02	0,07	0,09	0,01	0,04
	Δ_2 (%)	-8,96	-9,18	-8,52	-4,53	-4,67	-4,40
0.4	Δ_1(%)	0,82	0,14	0,03	0,41	0,07	0,01
	Δ_2 (%)	-6,28	-7,75	-7,67	-3,10	-3,90	-3,87
0.7	Δ_1 (%)	1,59	0,35	0,00	0,81	0,18	0,00
	Δ_2 (%)	-2,87	-5,91	-6,76	-1,26	-2,89	-3,35
1	Δ_1 (%)	2,51	0,62	0,06	1,31	0,32	0,03
	Δ_2 (%)	1,44	-3,55	-5,37	1,09	-1,60	-2,57
2	Δ_1 (%)	6,81	2,05	0,72	3,87	1,13	0,39
	Δ_2 (%)	25,34	9,49	3,57	14,76	5,87	2,55
3	Δ_1 (%)	12,39	4,02	1,85	8,16	2,49	1,10
	Δ_2 (%)	73,15	34,71	21,79	44,10	21,53	13,79

References

Akcay, A., Biller, B., and Tayur, S. (2011) Improved inventory targets in the presence of limited historical data. *Manufacturing Service Oper. Management* 13(3): 297-309.

Axsäter, S. (2015) *Inventory Control*, 3rd edition (Springer International, Switzerland).

Bookbinder, J.H. and Lordahl, A.E. (1989) Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Trans.* 21(4): 302-312.

Boylan, J.E. and Babai, M.Z. (2016) On the performance of overlapping and non-overlapping demand aggregation approaches. *Int. J. Prod. Econ.* 181, Part A: 136-144.

Boylan. J.E. and Syntetos, A.A. (2021) *Intermittent Demand Forecasting: Context, Methods and Applications.* (Wiley, Oxford).

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Ann. Stat.* 7(1): 1-26.

Fricker, R.D. and Goodhart, C.A. (2000) Applying a bootstrap approach for setting reorder points in military supply systems. *Nav. Res. Log.* 47(6): 459-478.

Gallego, G. and Moon, H. (1993) The distribution free newsboy problem: Review and extensions. *J. Opl. Res. Soc.* 44(8): 825-834.

Hasni, M., Babai, M.Z., Aguir, M.S. and Jemai, Z. (2019) An investigation on bootstrapping forecasting methods for intermittent demands. *Int. J. Prod. Econ.* 209, March 2019: 20-29.

Huh, W.T., Levi, R., Rusmevichientong, P., and Orlin, J.B. (2011) Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* 59(4): 929-941.

Lee, S., Kim, H., and Moon, I. (2021) A data-driven distributionally robust newsvendor model with a Wasserstein ambiguity set. *J. Opl. Res. Soc.* 72(8): 1879-1897.

Lengu, D., Syntetos, A.A., Babai, M.Z., (2013). Spare parts management: linking distributional assumptions to demand classification, *Eur. J. Oper. Res.*, 235(3): 624-635.

Perakis, G. and Roels, G. (2008) Regret in the newsvendor model with partial information. *Oper. Res.* 56(1): 188-203.

Porras, E. and Dekker, R. (2008) An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. *Eur. J. Oper. Res.* 184(1): 101-132.

Prak, D., Teunter, R.H., Babai, M.Z., Boylan, J.E., Syntetos, A.A. (2021). Robust compound Poisson parameter estimation for inventory control, *Omega*, 104: 102481.

Rego, J.R. and Mesquita, M.A. (2015) Demand forecasting and inventory control: A simulation study on automotive spare parts. *Int. J. Prod. Econ.* 161, March 2015: 1-16.

Rostami-Tabar, B., Babai, M.Z., Syntetos, A., and Ducq, Y. (2013) Demand forecasting by temporal aggregation. *Nav. Res. Log.* 60(6), 479-498.

Saghafian, S. and Tomlin, B. (2016) The newsvendor under demand ambiguity: combining data with moment and tail information. *Oper. Res.* 64(1): 167-185.

Siddiqui, R., Azmat, M., Ahmed, S., and Kummer, S. (2022) A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry. *Supply Chain Forum: An International Journal*, 23(2), 124-134.

Syntetos, A.A., Lengu, D., and Babai, M.Z. (2013) A note on the demand distribution of spare parts. *Int. J. Prod. Res.* 51: 6356-6358.

Turrini, L. and Meissner, J. (2019) Spare parts inventory management: New evidence from distribution fitting. *Eur. J. Opl. Res.* 273(1): 118-130.

Wang, M.C. and Rao, S.S. (1992) Estimating reorder points and other management science applications by bootstrap procedure. *Eur. J. Opl. Res.* 56(3): 332-342.

Willemain, T.R. and Smart, C.N. (2001) *System and Method for Forecasting Intermittent Demand*. United States Patent, Patent No. US 6,204,431 B1.

Willemain, T.R, Smart, C.N., Schockor, J.H. and DeSautels, P.A. (1994) Forecasting intermittent demand in manufacturing: A comparative evaluation of Croston's method. *Int. J. Forecasting* 10: 529-538.

Willemain, T.R, Smart, C.N. and Schwarz, H.F. (2004) A new approach to forecasting intermittent demand for service parts inventories. *Int. J. Forecasting* 20(3): 375-387.

Zhou, C. and Viswanathan, S. (2011) Comparison of a new bootstrapping method with parametric approaches for safety stock determination in spare parts inventory systems. *Int. J. Prod. Econ.* 133(1): 481-485.

Zhu, S., Dekker, R., van Jaarsveld, W., Renjie, R.W. and Koning, A.J. (2017) An improved method for forecasting spare parts demand using extreme value theory. *Eur. J. Oper. Res.* 261(1): 169-181.