Forecasting of lead-time demand variance: implications for safety stock calculations

Zied Babai, Yong Dai, Qinyun Li, Aris Syntetos, Xun Wang

PII: S0377-2217(21)00331-3
DOI: https://doi.org/10.1016/j.ejor.2021.04.017
Reference: EOR 17164

To appear in: European Journal of Operational Research

Received date: 11 July 2020
Accepted date: 8 April 2021

Please cite this article as: Zied Babai, Yong Dai, Qinyun Li, Aris Syntetos, Xun Wang, Forecasting of lead-time demand variance: implications for safety stock calculations, European Journal of Operational Research (2021), doi: https://doi.org/10.1016/j.ejor.2021.04.017

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.
Highlights

- We propose three strategies to estimate the variance of the lead-time demand
- Demand is assumed to be autocorrelated and lead-times are stochastic
- We derive analytical results under an ARMA(1,1) demand process
- We consider both SES and the MMSE forecasting method
- We show the underperformance of the classical strategy for positive autocorrelation
Forecasting of lead-time demand variance: implications for safety stock calculations

Zied Babai1,*; Yong Dai2; Qinyun Li3; Aris Syntetos3; Xun Wang3

1 Kedge Business School, France; 2 East China Normal University, PRC; 3 Cardiff University, UK

*Corresponding author:
Dr. M. Zied BABAI
Kedge Business School
680 cours de la Libération, 33405 Talence Cedex, France
Tel: +33 5 56 84 63 51, e-mail: mohamed-zied.babai@kedgebs.com

Abstract

Lead-time demand forecasting constitutes the backbone of inventory control. Although there has been a considerable amount of research on forecasting the mean lead-time demand, less has centered around forecasting lead-time demand variance, especially in the case of stochastic lead-times. This represents an important gap in the literature, given that safety stock calculations rely explicitly on the lead-time demand variance (or equivalently the variance of the lead-time demand forecast error for unbiased estimators). We bridge this gap by exploring the viability of three strategies to estimate the variance of the lead-time demand forecast error under stochastic lead-times: (1) aggregating the per period variance of forecast errors over the lead-time, which is the classical approach; (2) considering the variance of the aggregated (over the lead-time) forecast error; (3) considering the variance of the forecast errors resulting from temporally aggregated (over the lead-time length) demand. Analytical results are derived for a first order autoregressive moving average ARMA(1,1) demand process for both a single exponential smoothing and the minimum mean squared error forecasting method. A numerical investigation assesses the effects of demand autocorrelation and lead-time variability on the accuracy of each strategy, and the conditions under which one outperforms the others. The results show that the classical strategy presented in textbooks appears to be the least accurate one, except for cases with a high negative demand autocorrelation. An analysis of the inventory control performance also reveals that the classical strategy often leads to higher inventory costs and lower service levels for positive autocorrelation.

Keywords: Forecasting; Safety stocks; Demand variance; Auto-correlated demand; Lead-times
Introduction

Effective inventory management relies upon accurate lead-time demand forecasts (Graves, 1999; Prak et al., 2017). Undoubtedly, there has been a tremendous amount of research in the last decades in the area of demand forecasting (Syntetos et al., 2016a). However, most of these studies focus on the estimation of mean demand and very little work has centered around the estimation of the demand variance. The latter is equally important though since safety stock calculations are based explicitly on the lead-time demand variance – commonly expressed through the variance of the lead-time demand forecast error, when unbiased demand forecasts are considered (Gardner, 1988; Boute et al., 2014; Prak et al., 2017).

Some research has recently looked at the estimation of the variance of the lead-time demand by means of utilising the variance of the per-period forecast errors. However, this relies upon the assumption of i.i.d demand processes (Prak et al., 2017) and/or stationary auto-correlated demand series (Johnston and Harrison, 1986; Graves, 1999; Disney et al., 2006; Syntetos and Boylan, 2008) with constant lead-times. In the case of stationary i.i.d demand and constant lead-times, the classical approach suggested in most textbooks consists of multiplying the lead-time by the variance of the one-step ahead forecast error (Axsäter, 2006; Silver et al., 2017). Prak et al., (2017) have shown that such approach leads to an under-estimation of the lead-time demand variance, and consequently an under-estimation of the safety stocks needed to meet a target service level; this is because forecast errors are auto-correlated over the lead-time, although demand isn’t. Subsequently, they proposed appropriate mark-up adjustments for safety stock calculations. However, the corrected expression applies only to the case of a constant lead-time.

Alternative approaches that move beyond the per-period variance of forecast errors have also been considered in the literature. Syntetos and Boylan (2006) have looked at the aggregated forecast errors over a constant lead-time and their variance to calculate the variability of lead-time demand adopting a smoothed mean squared error (MSE) approach. Under the assumption of autocorrelated demand and a constant lead-time, Rostami-Tabar et al., (2013, 2014) have evaluated the variance of cumulative lead-time demand forecast error before and after temporal demand aggregation, where mean demand forecasts are generated based on single exponential smoothing (SES). Note that no stochastic lead-times are considered in the above described research works.

In this paper, we assess the viability of three distinct strategies to estimate the demand variance over a stochastic lead-time. The first strategy (Strategy 1) consists of aggregating the estimated variances of the per-period forecast errors. The second strategy (Strategy 2) consists of estimating the variance of
the cumulative lead-time demand forecast error. The third strategy (Strategy 3) is based on the temporal demand aggregation approach, i.e. the estimation of the variance of the demand forecast error is performed by considering the forecasts of the aggregated (over the lead-time) demand. Under each strategy, we derive the analytical expression of the variance of the lead-time demand forecast error. We do so for a first order autoregressive moving average, ARMA(1,1), demand process that is forecasted by means of single exponential smoothing; results under the minimum mean squared error forecasting method are also given in an electronic companion of the paper. The ARMA framework is chosen because in addition to its theoretical attractiveness (Box and Jenkins, 1970), empirical evidence supports its suitability in different real world supply chain settings, including retailing and machine tooling (Nahmias, 1993; Chopra and Meindl, 2007; Ali, et al., 2012). With regard to SES, this is a very popular forecasting method in industry (Gardner, 2006), is associated with very good empirical performance (see, e.g., the M1 competition results, Makridakis et al., 1982) and is an unbiased estimator in the case of ARMA(1,1) demands (Hsieh et al., 2020). Our analytical results are complemented by a numerical study in which we explore the effects of demand autocorrelation and lead-time variability on the performance of the three strategies.

We find that in most cases of autocorrelated demand, and under a stochastic lead-time, Strategies 2 and 3 often lead to the most accurate estimate of the lead-time demand variance along with the best inventory performance. The results also reveal that the classical strategy presented in textbooks (Strategy 1) appears to be accurate, mainly in cases associated with a negative demand autocorrelation. We also show that when the autocorrelation and moving average parameters of the demand are equal, the classical strategy presented in textbooks and the temporal aggregation strategy lead to the same estimate.

In summary, the contribution of our work is as follows:

1) We propose three strategies to estimate the variance of the lead-time demand forecast error, which can be used to calculate safety stocks, when the lead-time is stochastic and the demand is forecasted;
2) We derive the expressions of the variance estimates associated with the three proposed strategies under an ARMA(1,1) demand process and both a single exponential smoothing and the minimum mean squared error forecasting method;
3) We numerically analyse and compare the estimates of the three strategies as well as their inventory performance;
4) We show that, for high positive autocorrelation, the classical strategy presented in textbooks is the least accurate strategy that also leads to higher inventory costs.

Note that the generality of our models improves the explanatory efficacy. The results from established studies such as Axsäter et al., (2006), Rostami-Tabar et al., (2014), Prak et al., (2017), Silver et al.,
can all be derived from our models after simplification. More importantly, our models and results unfold the enigma in managing safety stock caused by the two types of stochasticity commonly seen – demand autocorrelation and lead-time variability. To the best of our knowledge, this is the first contribution revealing the general conditions under which one strategy of estimating the variance of lead-time demand forecast error outperforms others, with subsequent implications for inventory performance. This offers important managerial insights to practitioners.

The remainder of the paper is organised as follows. The relevant research is reviewed in the next section. In Section 3, we present the assumptions of our work and the theoretical developments for calculating the variance of forecast errors over a stochastic lead-time under the three strategies. In Section 4, the comparative performance of the three strategies is discussed under some particular cases of demand and insights are provided. Section 5 presents the results of a numerical analysis conducted to compare the performance of the three strategies and assess their sensitivity in terms of variance estimation and inventory performance. We conclude in Section 6 with what we believe are important insights for practical applications and worth-considering avenues for further research.

2. Research background

Most of the inventory control textbooks (e.g. Axsäter et al., 2006; Silver et al., 2017) indicate that in practical applications, and under a parametric forecasting approach, the parameters of the demand distribution (i.e. the mean and the variance of the demand, most commonly) have to be estimated, and that the resulting forecast error should be taken into account. The suggestion then is to use the variance of the per period demand forecast errors (e.g. obtained via the smoothed Mean Squared Error, MSE, or smoothed Mean Absolute Deviation, MAD, Brown, 1959; 1982) to determine safety stocks (by multiplying the lead-time by the variance of the one step ahead forecast error). It has also been shown that the smoothed MSE approach may be more efficient than the MAD one (Bretschneider, 1986).

Prak et al., (2017) have shown that even if demand is stationary, forecast errors are auto-correlated over the lead-time. Therefore, the process of calculating the variance of the lead-time demand forecast error by simply multiplying the lead-time by the per period variance of the forecast error is not correct. Syntetos and Boylan (2006) suggest a cumulative (smoothed) MSE procedure for directly forecasting the variance of demand over the lead-time through aggregating the per-period demand forecast errors. This suggestion can be regarded as a strategy of aggregation for variance estimation. By doing so, problems related to the relationship between the forecast errors over the lead-time are avoided. However, this cannot accommodate stochastic lead-times. In the case of auto-correlated demands, the literature contains already methods to estimate the variance of the lead-time demand (forecast error) under optimal and non-optimal forecasting methods (Johnston and Harrison, 1986; Graves, 1999; Lee
et al., 2000; Syntetos and Boylan, 2008). However, the relevant research has been developed under the rather constraining case of constant lead-times.

Although the assumption of constant lead-times introduces analytical convenience, there is considerable empirical evidence to suggest that lead-times are variable, reflecting erupted transportation schedules, capacity limitations etc. (Raff, 2006; Boute et al., 2007; Jaksic et al., 2011). (We refer here to the lead-times related to one particular Stock Keeping Unit, SKU, as opposed to the behavior of lead-times across SKUs.) Various continuous distributions have been assumed in previous research such as exponential (He et al., 2005; Bahri and Mohammad, 2012), normal (Bagchi et al., 1983; Hoque, 2013), gamma (Johansen and Thorstenson, 1993) and Erlang (Bagchi and Hayya, 1984; Kim et al., 2004; Johansen, 2005) to represent lead-times. Discrete lead-time distributions have also been considered, such as uniform (Chopra, 2004; Rao, 2005), Poisson (Suneung, 2008), negative binomial (Song et al., 2000), geometric (Lawrence et al., 2013) and arbitrary (Disney et al., 2016; Wang and Disney, 2017). Discrete lead-time distributions are better suited to the demand forecasting practice because demand data are often stored in discrete time buckets. Thus, is impermeable to estimate the lead-time demand when the lead-time is continuously distributed. In this paper, we relax the assumption of constant lead-times by considering discrete stochastic lead-times.

Furthermore, temporal demand aggregation is a forecasting approach based on which forecasting is attempted at some lower frequency time units than those where the data is originally recorded, and using that aggregated series for forecasting purposes (Nikolopoulos et al. 2011; Babai et al., 2012). For example, quarterly data can be obtained when aggregating monthly data in time-buckets of size three. Once the forecasts have been produced they need then to be disaggregated somehow to the original (monthly) time buckets. In an inventory context though, and in the case where the lead-time is actually three months (periods), no disaggregation is required. Temporal aggregation is an intuitively appealing approach to reduce demand uncertainty (Boylan and Babai, 2016). Assuming a stationary behavior, data in higher levels of aggregation (lower frequencies) are generally less volatile. One issue with temporal aggregation is the determination of the ‘optimal’ level of aggregation, a problem that becomes redundant though (as discussed above) when the aggregation level is set equal to the forecast horizon reflecting the context of application (i.e. lead-time in an inventory setting). This simple approach may bring gains in terms of forecast accuracy (Nikolopoulos et al., 2011) and inventory performance (Babai et al., 2012). Rostami-Tabar et al., (2013, 2014) have compared the MSE of the forecast before and after aggregation at the original and aggregated levels (under the assumption of a constant lead-time). They found that temporal aggregation is a useful approach in order to reduce the variance of the forecast error. This approach is also considered in our work to estimate the variance of the lead-time demand forecast error under stochastic lead-times.

3. Strategies for forecast error variance calculation
We use the following notation throughout the paper.

\( n \): Sample size, the length of the original time series
\( t \): Time period in the original time series, \( t = 1, 2, \ldots, n \)
\( d_t \): Demand at period \( t \)
\( \mu_d \): Mean demand per period
\( \sigma_d^2 \): Variance of demand per period
\( \gamma_k \): Covariance of lag \( k \) demands, \( \gamma_k = \text{cov}(d_t, d_{t+k}) \)
\( cv_d \): The coefficient of variation of the demand, \( cv_d = \sigma_d/\mu_d \).
\( f_{t+i} \): Forecast for demand in period \( t + i \) (\( d_{t+i} \)), made at the end of period \( t \), \( i = 1, 2, \ldots, L \)
\( \mu_f \): Mean demand forecast per period
\( \sigma_f^2 \): Variance of demand forecast per period
\( L \): Lead-time
\( \mu_L \): Mean lead-time
\( \sigma_L^2 \): Variance of lead-time
\( D^L \): Demand over the lead-time
\( F^L \): Forecast of the lead-time demand
\( T \): Time period in the aggregated time series, \( T = 1, 2, \ldots, \left[ \frac{n}{L} \right] \)

\( \text{LTFE} \): Lead-time (demand) forecast error
\( \epsilon_{t+i} \): Error of the forecast made at time \( t \) for the demand in period \( t+i \)
\( \phi \): Autoregressive parameter for the original time series, \( 0 < |\phi| < 1 \)
\( \phi' \): Autoregressive parameter for the time series after temporal aggregation, \( 0 < |\phi'| < 1 \)
\( \theta \): Moving average parameter for the original time series, \( 0 < |\theta| < 1 \)
\( \alpha, \beta \): Smoothing constants for SES forecasting, \( 0 \leq \alpha, \beta \leq 1 \)

\( \text{Prob}(L = l) \): The probability that the lead-time \( L \) equals \( l \)
\( \text{E}[x] \): Expected value of \( x \).

We assume stationary demand following an autoregressive moving average process of order 1, \( \text{ARMA}(1,1) \). As such, the demand at any period \( t \) can be expressed as

\[
d_t = \mu + \phi d_{t-1} + \epsilon_t - \theta \epsilon_{t-1}, \quad \epsilon_t \sim (0, \sigma^2).
\]

For such (demand) series, it is known that:

\[
\mu_d = \frac{\mu}{1 - \phi}
\]

\[
\gamma_0 = \sigma_d^2 = \frac{1 - 2\phi \theta + \theta^2}{1 - \phi^2} - \sigma^2
\]

\[
\gamma_1 = \text{cov}(d_t, d_{t+1}) = \frac{(1 - \phi \theta)(\phi - \theta)}{1 - \phi^2} \sigma^2
\]
\( y_k = \text{cov}(d_t, d_{t+k}) = \phi^{k-1} y_1 \quad \text{for any } k \geq 2 \)  

We also assume that SES is used to forecast the demand at any future time period.

In the following subsections, we derive the expression of the variance of the lead-time demand forecast error under the three considered strategies, followed by some details about their practical implementation. The expressions where the minimum mean squared error forecasting method is used are presented in the electronic companion of the paper. Note that for presentation purposes and since we use an unbiased forecast method, the terms lead-time demand variance and the variance of lead-time demand forecast error are used interchangeably.

### 3.1. Strategy 1: Aggregation of the per period variances of forecast errors

Strategy 1 constitutes an extension of the classical approach where the variance of lead-time demand forecast error is estimated as the variance of the per-period forecast error multiplied by the lead-time (Silver et al., 2017). We rather take the sum of the variances of forecast errors, i.e. \( \sum_{i=1}^{l} \text{var}(e_{t+i}) \) as our demand process is auto-correlated.

The variance of lead-time demand forecast error, denoted by \( \text{var}(\text{LTFE})_{S_1} \), can be written as:

\[
\text{var}(\text{LTFE})_{S_1} = E[\text{var}(\text{LTFE}|L = l)] + \text{var}(E[\text{LTFE}|L = l])
\]

We first calculate the conditional variance of forecast error over the lead-time when the lead-time \( L \) takes on a given value of \( l \), \( \text{var}(\text{LTFE}|L = l) \), which can be expressed as:

\[
\text{var}(\text{LTFE}|L = l) = \text{var}(e_{t+1}) + \text{var}(e_{t+2}) + \cdots + \text{var}(e_{t+l})
\]

\[
= \text{var}(d_{t+1} - f_{t+1}) + \text{var}(d_{t+2} - f_{t+1}) + \cdots + \text{var}(d_{t+l} - f_{t+1})
\]

\[
= \sigma_d^2 + l\sigma_f^2 - 2(\text{cov}(d_{t+1}, f_{t+1}) + \cdots + \text{cov}(d_{t+l}, f_{t+1}))
\]

Under SES forecasting with a smoothing constant \( \alpha \), the variance of the forecasts and the covariance between \( d_{t+i} \) and \( f_{t+i} \) are

\[
\sigma_d^2 = \frac{\alpha}{2 - \alpha} \sigma_d^2 + \frac{2\alpha(1 - \alpha)}{(2 - \alpha)(1 - \phi + \alpha\phi)} y_1
\]

\[
\text{cov}(d_{t+i}, f_{t+i}) = \frac{\alpha\phi^{i-1} y_1}{1 - \phi + \alpha\phi}
\]

Substituting (7) and (8) into (6), we get

\[
\text{var}(\text{LTFE}|L = l) = l \left(1 + \frac{\alpha}{2 - \alpha}\right) \sigma_d^2 + \frac{2\alpha}{1 - \phi + \alpha\phi} \left(\frac{1 - \phi}{1 - \phi - \alpha l} - 1 - \phi l\right) y_1
\]

As SES forecasts are unbiased, the conditional mean of the forecast error over the lead-time is \( E[\text{LTFE}|L = l] = 0 \). Then, the variance of lead-time demand forecast error under Strategy 1 can be written as:

\[
\text{var}(\text{LTFE})_{S_1} = E[\text{var}(\text{LTFE}|L = l)] + \text{var}(E[\text{LTFE}|L = l])
\]

\[
= E[\text{var}(e_{t+1}) + \text{var}(e_{t+2}) + \cdots + \text{var}(e_{t+l})]
\]
\[ \sum \operatorname{Prob}(L = l) \left[ l \left( 1 + \frac{\alpha}{2 - \alpha} \right) \sigma_d^2 + \frac{2\alpha}{1 - \varnothing + \alpha \varnothing} \left( \frac{1 - \alpha}{2 - \alpha} \right) \sigma_a \right] \]
\[ = \frac{2}{2 - \alpha} \mu_L \sigma_d^2 + \frac{2\gamma_1}{1 - \varnothing + \alpha \varnothing} \left( \mu_L (1 - \alpha) \right) - E(1 - \varnothing) \]  
\[ \tag{10} \]

### 3.2. Strategy 2: Variance of the aggregated forecast error

Under Strategy 2, the variance is estimated by calculating the variance of the cumulative lead-time demand forecast error, i.e. \( \text{var} \left( \sum_{i=1}^{L} e_{t+i} \right) \). This is different from Strategy 1 in that the correlation of forecast errors rather than of the demands (e.g., Prak et al., 2017) is (implicitly) taken into account. The variance of the lead-time demand forecast error under Strategy 2, denoted by \( \text{var}(LTFE)_{S2} \), is given by:

\[ \text{var}(LTFE)_{S2} = \text{var} \left( \sum_{i=1}^{L} e_{t+i} \right) = \text{var}(D^L - F^L) \]
\[ = \text{var}(D^L) + \text{var}(F^L) - 2\text{cov}(D^L, F^L) \]  
\[ \tag{11} \]

We analyse each term in (11) separately. The lead-time demand variance is:

\[ \text{var}(D^L) = E[\text{var}(D^L|L = l)] + \text{var}[E(D^L|L = l)] \]
\[ = \sum_l \operatorname{Prob}(L = l) \text{var}(d_{t+1} + d_{t+2} + \cdots + d_{t+L}) + \text{var}(\mu_d) \]
\[ = \sum_l \operatorname{Prob}(L = l) \left( l \sigma_d^2 + 2\text{cov}(d_{t+1}, d_{t+2}, \cdots, d_{t+L}) + \sigma_d^2 \right) + \sigma_d^2 \mu_d^2 \]
\[ = \mu_L \sigma_d^2 + 2\gamma_1 \left( \frac{l - 1 - \varnothing - \varnothing^L}{1 - \varnothing} \right) \gamma_1 + \mu_L^2 \sigma_d^2 \]  
\[ \tag{12} \]

Under SES, the demand forecast over the lead-time is \( F^L = \sum_{i=1}^{L} f_{t+i} = L f_{t+1} \). The variance of the lead-time demand forecast is

\[ \text{var}(F^L) = E[\text{var}(F^L|L = l)] + \text{var}[E(F^L|L = l)] \]
\[ = E[\text{var}(L f_{t+1})] + \text{var}[E(L f_{t+1})] \]
\[ = E[l^2 \text{var}(f_{t+1})] + \text{var}(l E[f_{t+1}]) \]
\[ = E[l^2 \sigma_f^2] + \text{var}(l \mu_f) \]

which leads to

\[ \text{var}(F^L) = \sigma_f^2 (\mu_f^2 + \sigma_f^2) + \mu_f^2 \sigma_f^2 \]  
\[ \tag{13} \]

The covariance is obtained as follows:
$$\text{cov}(D^L, F^L) = E[\text{cov}(D^L, F^L)|L = l] + \text{cov}(E[D^L|L = l], E[F^L|L = l])$$

$$= \sum_l Pr(L = l) \text{cov} \left( \sum_{i=1}^l d_{t+i}, l f_{t+1} \right) + \text{cov}(l \mu_d, l \mu_f)$$

$$= \sum_l Pr(L = l) l \text{cov} \left( \sum_{i=1}^l d_{t+i}, f_{t+1} \right) + \mu_d \mu_f \sigma_L^2$$

$$= \sum_l Pr(L = l) \frac{(1 - \theta^l)}{(1 - \phi)(1 - \phi + \alpha \phi)} \alpha \gamma_1 l + \mu_d \mu_f \sigma_L^2$$

$$= \frac{\alpha \gamma_1}{(1 - \phi)(1 - \phi + \alpha \phi)} E[L(1 - \theta^L)] + \mu_d \mu_f \sigma_L^2$$  \hfill (14)

Since for $\text{SES}, \mu_f = \mu_d$, and after substituting (14), (13) and (12) into (11), we have:

$$\text{var}(\text{LTFE})_{s2} = \mu_d \sigma_d^2 + \sigma_f^2 \left( \mu_d^2 + \sigma_f^2 \right) + \frac{2 \gamma_1 \mu_d (1 - \phi) - 1 + E[\theta^L]}{(1 - \phi)^2} \frac{2 \alpha \gamma_1 E[L(1 - \theta^L)]}{(1 - \phi)(1 - \phi + \alpha \phi)}$$  \hfill (15)

It should be noted that the estimate of $\text{var}(\text{LTFE})_{s2}$ given in (15) reduces, for constant lead-times (i.e., $L = \mu_L$), to the expression of the variance estimate (equation 8) in Rostami-Tabar et al., (2014).

It also reduces, for constant lead-times and i.i.d. demand, to the expression related to the third approach in Prak et al., (2017). Hence, the expression given in (15) is a generalisation of the expressions presented by Rostami-Tabar et al., (2014) and Prak et al., (2017).

### 3.3. Strategy 3: Variance of the forecast error by temporal aggregation

In Strategy 3, the variance calculation relies upon the errors associated with the aggregated (over the lead-time) mean demand forecast. Strategy 3 consists of two steps: i) first, the demand is temporally aggregated – with the aggregation level being equal to the mean lead-time; ii) then demand forecasting and variance estimation are carried out considering the aggregated series.

The SES forecasts of the temporally aggregated demand (with a smoothing constant $\beta$), are produced as follows:

$$F_{T+1} = \beta D_T + (1 - \beta) F_T.$$  

It is known that temporal aggregation of ARMA(1,1) processes leads to a (new) ARMA(1,1) process (Wei, 2006; Rostami-Tabar et al., 2013). The relationship between the original time series and the aggregated ones are as follows:

$$\gamma_0' = \text{var}(D_T|L = l) = \begin{cases} 
\gamma_0 & l = 1 \\
\gamma_0 + \gamma_1 \sum_{k=1}^{l-1} 2(l-k)\theta^{k-1} & l > 1
\end{cases}$$  \hfill (16)

$$\gamma_1' = \text{cov}(D_T, D_{T+1}|L = l) = \begin{cases} 
\gamma_1 & l = 1 \\
\gamma_1 \left( \sum_{k=1}^{l} k \theta^{k-1} + \sum_{k=2}^{l} (k-1)\theta^{2l-k} \right) & l > 1
\end{cases}$$  \hfill (17)
\[ \varphi' = \varphi_i \]  

The forecast variance and the covariance between the demand and the forecast at the aggregated series level are as follows:

\[ \text{var}(F_{t+1}|L = l) = \frac{\beta y_0'}{2 - \beta} + \frac{2\beta (1 - \beta) y_1'}{(2 - \beta)(1 - \varphi' + \beta \varphi')} \]  

\[ \text{cov}(D_{T+1}, F_{T+1}|L = l) = \frac{\beta y_1'}{1 - \varphi' + \beta \varphi'} \]

Under Strategy 3, the variance of the lead-time demand forecast error, denoted by \( \text{var}(LTFE)_{S3} \), can be written as:

\[ \text{var}(LTFE)_{S3} = E[\text{var}(D_{T+1} - F_{T+1})|L = l] + \text{var}[E(\text{var}(D_{T+1} - F_{T+1})|L = l)] \]  

Since the SES forecasts are unbiased for the ARMA(1,1) demand process, we have: \( \text{var}(E[(D_{T+1} - F_{T+1})|L = l]) = 0 \).

Using (16) and (17), \( \text{var}(LTFE)_{S3} \) is calculated as:

\[ \text{var}(LTFE)_{S3} = E[\text{var}(D_{T+1} - F_{T+1})|L = l] \]

\[ = \sum_{l=1}^{\infty} \text{Prob}(L = l) \left( \frac{2\sigma_d^2}{2 - \beta} - \frac{2\beta y_1}{(2 - \beta)(1 - \varphi' + \beta \varphi')} \right) + \sum_{l \in \omega} \text{Prob}(L = l) \left( \frac{2\sigma_d^2}{2 - \beta} - \frac{2\beta y_1}{(2 - \beta)(1 - \varphi' + \beta \varphi')} \right) \]

\[ = 2\sigma^2 \left( \frac{2 + \beta - \varphi' (4 - (2 - \beta) \varphi^l) - 2l(1 - \varphi) (1 - (1 - \beta) \varphi^l)}{(2 - \beta)(1 - \varphi^2)(1 - (1 - \beta) \varphi^l)} \right) \]

where \( \omega \) is denoted as a set \( \{2, 3, \ldots, n\} \).

It is easy to show that, for a constant lead-time (i.e., \( L = \mu_L \)), (22) reduces to

\[ \text{var}(LTFE)_{S3} = \frac{2L\sigma^2}{(2 - \beta)(1 - \varphi^2)} \left( 1 + \frac{\theta^2}{2} - \frac{2\theta \varphi}{1 - \varphi^2} \right) \]

\[ - \frac{2\sigma^2}{(2 - \beta)(1 - \varphi^2)} \left( 1 - \varphi \right) \left( 2 + \beta - 2L(1 - \varphi)(1 - (1 - \beta) \varphi^l) - \varphi^l (4 - (2 - \beta) \varphi^l) \right) \]

which can be shown to be equal to

\[ \text{var}(LTFE)_{S3} = 2\sigma^2 \left[ \frac{L(1 + \theta^2 - 2\theta \varphi)}{(2 - \beta)(1 - \varphi^2)} - \frac{2(-\theta + \varphi)(1 - \theta \varphi)(1 - L + L\varphi - \varphi^L)}{(2 - \beta)(1 + \varphi^2)(1 - \varphi^2)} \right] \]

\[ - \frac{2\beta(-\theta + \varphi)(1 - \theta \varphi)}{(2 - \beta)(1 - \varphi^2)} \left( \frac{\varphi^L (1 - L + L\varphi - \varphi^L)}{(1 + \varphi^2)} + \frac{1 - \varphi^L - L\varphi^L + L\varphi^{1+L}}{(1 + \varphi^2)} \right) \]

which is the same expression of the variance estimate (equation 10) in Rostami-Tabar et al., (2014). Therefore, the estimate of \( \text{var}(LTFE)_{S3} \) reduces, for constant lead-times, to the variance estimate under the aggregation approach analysed in Rostami-Tabar et al., (2014).
3.4. Practical implementation

From an implementation perspective, the problem of calculating the mean squared error (MSE) under a stochastic lead-time for each strategy (equivalently the safety stock) is similar to the case of a constant lead-time. In fact, the MSE is initialised using the expressions derived earlier in this section (expressions of the variance at the steady state using the average values) and then at any period \( t \), the MSE can be updated (using exponential smoothing for example) by considering the actual forecast error (actual observed forecast error over the actual lead-time). This follows the same methodology used in the literature when the lead-time is constant and the demand (assumed stationary) is forecasted (e.g. Syntetos and Boylan, 2008; Hasni et al., 2019).

4. Theoretical properties of the strategies

Due to the complexity related to the analytical comparison of the variance expressions derived under the three strategies in the general case of auto-correlated demands and stochastic lead-times, we first derive, in this section, some properties under several particular cases when SES is used. We then present the results of a numerical investigation in Section 5 where the performance of the strategies is considered in a broader range of settings. The properties and numerical results, when the MMSE forecasting method is used, are presented in the electronic companion of the paper. Note that a better performance of a particular strategy is reflected by a more accurate estimation of the lead-time demand variance, which in turn leads to a more accurate calculation of the safety stocks.

4.1 Case of i.i.d demand

It is important to note that Strategy 1 may be interpreted as an aggregation of the per period variances of forecast error, identical to the classical approach in many textbooks. This can be easily proved as follow:

For an i.i.d. demand, i.e., \( \Theta = \theta = 0 \), (10) reduces to

\[
\text{var}(\text{LTFE})_{S1} = \mu_L \sigma_d^2 \frac{2}{2 - \alpha}
\]

This can also be written as

\[
\text{var}(\text{LTFE})_{S1} = \mu_L \sigma_d^2 + \mu_L \sigma_f^2
\]

We know that (e.g. Silver et al., 2017)

\[
\text{Lvar}(e_{t+1}) = \text{Lvar}(d_{t+1} - f_{t+1}) = \mu_d^2 + \sigma_f^2
\]

By letting \( L = \mu_L \), it is clear that (25) is identical to (24).

Furthermore, it is easy to show that when \( \theta = 0 \) (\( \mu_f = \mu_d \)), Strategy 1 leads to \( \text{var}(\text{LTFE})_{S1} = \mu_L \sigma_d^2 \) which is identical to the lead-time demand variance expression for constant lead-times. If the smoothing parameter value increases, the variance of forecasts increases, resulting in an increase in
the variance of the lead-time demand forecast error under Strategy 1. It is intuitively appealing that a larger safety stock is needed when forecast accuracy deteriorates.

For $\theta = \theta = 0$ with a constant lead-time, (22) reduces to

$$\text{var}(LTFE)_{S3} = \mu_L \sigma_d^2 \frac{2}{2 - \beta}$$

which is identical to (23) (i.e. Strategy 1 and Strategy 3 are similar) if the same smoothing parameter $\alpha = \beta$ is used for SES. This is due to the fact that the covariance of the demand (before aggregation) and forecast under an i.i.d. demand is zero. Therefore, the variance of the forecast error after temporal aggregation is only affected by the demand variance and the forecast variance over the lead-time after aggregation, which implies the same expression, compared with Strategy 1. However, it is important to note though that the exponential smoothing parameter $\beta$ at the aggregated series should often be different to the smoothing parameter $\alpha$ on the original series (Rostami-Tabar et al., 2013).

**Proposition 1.** The estimated variance of the lead-time demand forecast error resulting from Strategy 1 is lower than that resulting from Strategy 2.

**Proof.**

If there is no demand autocorrelation, the unconditional variance of lead-time demand (12) and the unconditional variance of lead-time demand forecast (13) in Strategy 2 reduce to (26) and (27) respectively.

$$\text{var}(D^L) = \mu_L \sigma_d^2 + \mu_L^2 \sigma_L^2$$

$$\text{var}(F^L) = \frac{\alpha}{2 - \alpha} (\mu_L^2 + \sigma_d^2) \sigma_d^2 + \mu_L^2 \sigma_L^2$$

Note that (26) is identical to the expression given in standard inventory textbooks.

When an unbiased forecasting method is used, the estimation of the lead-time demand variance is conducted through the variance of the lead-time demand forecast error. The unconditional covariance can be rewritten as

$$\text{cov}(D^L, F^L) = \mu_L \sigma_d^2$$

Under Strategy 2, (15) is simplified to:

$$\text{var}(LTFE)_{S2} = \mu_L \sigma_d^2 + \frac{\alpha}{2 - \alpha} (\mu_L^2 + \sigma_d^2) \sigma_d^2 = \mu_L \sigma_d^2 + \mu_L^2 \sigma_L^2$$

Based on (11), it is easy to observe that $\mu_L^2 \sigma_L^2$ in $\text{var}(D^L)$ and the covariance between the lead-time demand and the lead-time forecast (27) cancel each other out. Thus, the effect of $\mu_L^2 \sigma_L^2$ is eliminated from the variance estimation of the lead-time demand forecast error.

Comparing (29) to (24), whenever the lead-time variability $\sigma_L^2$ is considered, $\mu_L^2 \sigma_d^2 + \sigma_L^2 \sigma_d^2$ is strictly larger than $\mu_L \sigma_d^2$ when $\sigma_d^2 \neq 0$ and $\mu_L > 1$. As a result, we show that $\text{var}(LTFE)_{S2} > \text{var}(LTFE)_{S1}$.

The observations made by Prak et al. (2017) are similar to what we show now in the case of stochastic
lead-times. The difference between $\text{var}(LTFE)_{S2}$ and $\text{var}(LTFE)_{S1}$ increases with the variability of the lead-times.

\begin{proof}

Proposition 2. Strategy 1 underestimates the lead-time demand variance when

$$cv_d^2 < \frac{(2 - \alpha)\sigma_d^2}{\alpha \mu_t}$$

whereas Strategy 2 underestimates the lead-time demand variance when

$$cv_d^2 < \frac{2 - \alpha}{\alpha} \left( \frac{\sigma_d^2}{\mu_t^2 + \sigma_t^2} \right)$$

\end{proof}

The difference between the variance of Strategy 1 given by (10) and the lead-time demand variance (26) can be simplified to

$$\Delta_{S1} = \text{var}(LTFE)_{S1} - \text{var}(D^{L}) = \mu_t\sigma_f^2 - \mu_d^2\sigma_t^2 = \frac{\alpha}{2 - \alpha}\mu_t\sigma_d^2 - \mu_d^2\sigma_t^2$$

The difference between (29) and (26) is

$$\Delta_{S2} = \text{var}(LTFE)_{S2} - \text{var}(D^{L}) = (\mu_t^2 + \sigma_t^2)\sigma_f^2 - \mu^2\sigma_t^2 = \frac{\alpha}{2 - \alpha}(\mu_t^2 + \sigma_t^2)\sigma_d^2 - \mu_d^2\sigma_t^2$$

$\Delta_{S1} < 0$ means that Strategy 1 underestimates the lead-time demand variance. This holds for

$$cv_d^2 < \frac{(2 - \alpha)\sigma_d^2}{\alpha \mu_t}$$

Whereas, we have $\Delta_{S2} < 0$ when

$$cv_d^2 < \frac{2 - \alpha}{\alpha} \left( \frac{\sigma_d^2}{\mu_t^2 + \sigma_t^2} \right)$$

which proves Proposition 2.

The estimation performance of Strategy 1 and Strategy 2 is determined by the forecast variance (which is a function of $\alpha$), the coefficient of variation of demand ($cv_d$), and lead-time variability. Strategy 1 and Strategy 2 always underestimate the lead-time demand variance when the forecasts are less volatile (i.e. for low smoothing constant $\alpha$), a highly variable lead-time or when $cv_d$ is small. However, Strategy 1 and Strategy 2 may lead to an overestimation of the lead-time demand variance for a lead-time with very low variability and high demand volatility.

Combining Proposition 1 and Proposition 2, we can deduce Proposition 3.

\begin{proof}

Proposition 3. Strategy 2 better estimates the lead-time demand variance than Strategy 1 if

$$cv_d^2 < \frac{2 - \alpha}{\alpha} \left( \frac{\sigma_d^2}{\mu_t^2 + \sigma_t^2} \right)$$

\end{proof}
\[ cv_d^2 \in \left( \frac{2 - \alpha}{\alpha} \left( \frac{\sigma_L^2}{\mu_L^2 + \sigma_L^2} \right), \frac{(2 - \alpha)\sigma_L^2}{\alpha \mu_L} \right) \]

and this is reversed if

\[ cv_d^2 \in \left( \frac{2 - \alpha}{\alpha} \left( \frac{2\sigma_L^2}{\mu_L^2 + \mu_L + \sigma_L^2} \right), \frac{(2 - \alpha)\sigma_L^2}{\alpha \mu_L} \right) \]

Proof.

Based on (30) and (31) and the fact that in practical inventory settings with stochastic lead-times \( \mu_L^2 + \sigma_L^2 > \mu_L \) we have

\[ \frac{2 - \alpha}{\alpha} \left( \frac{\sigma_L^2}{\mu_L^2 + \sigma_L^2} \right) < \frac{(2 - \alpha)\sigma_L^2}{\alpha \mu_L} \] (32)

It is easy to see that (31) is a sufficient condition for both Strategy 1 and Strategy 2 to underestimate the lead-time demand variance. In addition, if Strategy 2 underestimates the lead-time demand variance, Strategy 1 does so too.

Furthermore, if (30) does not hold, it is a sufficient condition that both strategies 1 and 2 overestimate the lead-time demand variance, i.e. if Strategy 1 overestimates it, Strategy 2 also does. Hence, we can conclude that when

\[ cv_d^2 \leq \frac{2 - \alpha}{\alpha} \left( \frac{\sigma_L^2}{\mu_L^2 + \sigma_L^2} \right) \] (33)

Strategy 2 leads to a more accurate estimate of the lead-time demand variance than Strategy 1, whereas this is reversed when

\[ cv_d^2 \geq \frac{(2 - \alpha)\sigma_L^2}{\alpha \mu_L} \] (34)

We now study the situation when Strategy 1 underestimates the variance of the lead-time demand (i.e. (30) holds) and Strategy 2 overestimates the variance of the lead-time demand (i.e. (31) does not hold). Solving \(-\Delta_{s1} > \Delta_{s2}\), we get

\[ cv_d^2 > \frac{2 - \alpha}{\alpha} \left( \frac{2\sigma_L^2}{\mu_L^2 + \mu_L + \sigma_L^2} \right) \] (35)

It is easy to prove that

\[ \frac{\sigma_L^2}{\mu_L^2 + \sigma_L^2} < \frac{2\sigma_L^2}{\mu_L^2 + \mu_L + \sigma_L^2} < \frac{\sigma_L^2}{\mu_L} \] (36)

Thus, the situation where (30) and (35) hold while (31) does not hold occurs when

\[ \frac{2 - \alpha}{\alpha} \left( \frac{2\sigma_L^2}{\mu_L^2 + \mu_L + \sigma_L^2} \right) < cv_d^2 < \frac{(2 - \alpha)\sigma_L^2}{\alpha \mu_L} \] (37)

Inequality (37) offers another sufficient condition for Strategy 2 to give a more accurate estimate of the lead-time demand variance.
When $CV_d^2 < \frac{2-\alpha}{\alpha} \left( \frac{2\sigma_L^2}{\mu_L^2 + \mu_L + \sigma_L^2} \right)$, we have $-\Delta_{S1} < \Delta_{S2}$.

Based on (36), it is clear that (30) still holds in this case. Considering the condition that (30) does not hold, Strategy 1 provides a more accurate estimate of the lead-time demand variance when the condition given by (34) holds.

$$\frac{2-\alpha}{\alpha} \left( \frac{\sigma_L^2}{\mu_L^2 + \sigma_L^2} \right) < CV_d^2 < \frac{2-\alpha}{\alpha} \left( \frac{2\sigma_L^2}{\mu_L^2 + \mu_L + \sigma_L^2} \right)$$  (38)

This ends the proof of Proposition 3

Propositions 1-3 offer important advice on choosing between Strategy 1 and Strategy 2. In addition, Proposition 2 suggests the conditions of underestimating the lead-time demand variance, which has direct implications for safety stock calculations.

Therefore, we provide the following result given in Proposition 4.

**Proposition 4.** For high volume i.i.d. demand, the classic estimation $\mu_L \sigma_d^2 + \mu_L \sigma_f^2$ should be modified to $\mu_L \sigma_d^2 + (\mu_L^2 + \sigma_L^2) \sigma_f^2$, if taking into account the lead-time variability. The latter better approximates the lead-time demand variance than the former.

In the modified expression of lead-time demand variance estimation, the existence of lead-time variability is asserted by $(\mu_L^2 + \sigma_L^2) \sigma_f^2$. It also indicates that the impact of lead-time variability on the variance estimation is influenced by the forecast variability, $\sigma_f^2$. When $\alpha = 0$, $\sigma_f^2 = 0$, thus the estimation of lead-time demand variance via Strategy 2 is $\mu_L \sigma_d^2$ regardless of the lead-time variability. Nevertheless, when $\alpha \neq 0$, $\sigma_f^2 \neq 0$, then if the variance of forecasts increases the impact of $\mu_L^2 + \sigma_L^2$ is magnified, and the estimation of lead-time demand variance also increases, although the lead-time distribution does not change.

### 4.2 Case of auto-correlated demand

By simplifying (12), (10), (15), and (22), it is not difficult to prove that for stationary auto-correlated demand with constant lead-times, if $\alpha, \beta \to 0$, Strategy 2 and Strategy 3 exactly estimate the lead-time demand variance, while Strategy 1 is underestimating it for $\emptyset > \theta$ or overestimating it for $\emptyset < \theta$ if the lead-time is greater than one.

**Proposition 5.** If $\alpha, \beta \to 0$, neither demand autocorrelation nor the lead-time stochasticity is reflected in Strategy 1’s estimate. Strategy 2 and Strategy 3 reduce to the same strategy which leads to a higher variance estimation of the lead-time demand forecast error as compared to Strategy 1 when $\emptyset > \theta$, and a lower one when $\emptyset < \theta$.

Proof.
When $\alpha, \beta \to 0$, the variance of the lead-time demand under Strategy 1, 2 and 3 is given by (39), (40) and (41), respectively.

$$\text{var}(LTFE)_{S1} = \mu_L \sigma_d^2$$  \hspace{1cm} (39)

$$\text{var}(LTFE)_{S2} = \mu_L \sigma_d^2 + \frac{2\gamma_1 (\mu_L (1 - \phi) - 1 + E[\phi^L])}{(1 - \phi)^2}$$  \hspace{1cm} (40)

$$\text{var}(LTFE)_{S3} = \mu_L \sigma_d^2 + \frac{2\gamma_1 (\mu_L (1 - \phi) - 1 + E[\phi^L])}{(1 - \phi)^2}$$  \hspace{1cm} (41)

Eq. (39) clearly shows that neither the demand autocorrelation (and consequently the autocorrelation of the forecast errors) nor the lead-time variability is reflected in the estimate of the variance of lead-time demand forecast error.

It is also clear that (40) and (41) are the same and thus Strategy 2 and Strategy 3 reduce to the same strategy. The difference between (39) and (40) can be easily shown to be equal to:

$$\text{var}(LTFE)_{S2} - \text{var}(LTFE)_{S1} = \frac{2\gamma_1 (\mu_L (1 - \phi) - 1 + E[\phi^L])}{(1 - \phi)^2}$$

For $L = 1, L(1 - \phi) - 1 + \phi^L = 0$; and

$$L(1 - \phi) - 1 + \phi^L > 0 \forall L > 1$$  \hspace{1cm} (42)

Then we have

$$\mu_L (1 - \phi) + E[\phi^L] \geq 0$$

which holds for any stochastic discrete lead-time distribution.

Therefore, the difference between $\text{var}(LTFE)_{S2}$ and $\text{var}(LTFE)_{S1}$ has the same sign as $\gamma_1$. If $\phi > \theta$, $\gamma_1$ is positive, and then the difference is positive. When $\phi < \theta$, $\gamma_1$ is negative, and the difference is negative. \hfill \Box

Considering Proposition 5 and the lead-time demand variance expressed in (12), we provide Propositions 6 and 7.

**Proposition 6.** When $\alpha, \beta \to 0$, Strategy 2 and Strategy 3 always underestimate the lead-time demand variance while there is a threshold $T$ to determine Strategy 1’s (over/under) performance.

**Proof.**

The difference between (39) and (12) is

$$\Delta_{S1} = \text{var}(LTFE)_{S1} - \text{var}(D^L) = -\sigma_d^2 \mu_d^2 - \frac{2\gamma_1 (\mu_L (1 - \phi) - 1 + E[\phi^L])}{(1 - \phi)^2}$$  \hspace{1cm} (43)

Proposition 5 shows that Strategy 2 and 3 are the same when $\alpha, \beta \to 0$, then the difference between Strategy 2 (or equivalently Strategy 3) and (12) can be denoted as

$$\Delta_{S2} = \text{var}(LTFE)_{S2} - \text{var}(D^L) = -\sigma_d^2 \mu_d^2$$  \hspace{1cm} (44)
\(\Delta_{s_2}\) is always negative, indicating that Strategy 2 and Strategy 3 always underestimate the lead-time demand variance.

\(\Delta_{s_1}\) can be positive, negative and zero. Solving \(\Delta_{s_1} = 0\), we have

\[
cv_d^2 = -\frac{(1 - \theta)^2(1 - 2\theta \phi + \theta^2)\sigma_i^2}{2(1 - \theta)(\phi - \theta)(\mu_i(1 - \phi) - 1 + E[\phi^2])} = T
\]

\(T\) denotes the threshold value that determines over- and under-estimation resulting from Strategy 1.

When \(cv_d^2 < T\), Strategy 1 underestimates the lead-time demand variance; when \(cv_d^2 > T\), it overestimates it.

Proposition 7. When \(\alpha, \beta \to 0\), Strategies 2 and 3 better approximate the lead-time demand variance than Strategy 1 when

\[
\phi > \theta \text{ or } \begin{cases} 
\phi < \theta \\
\frac{cv_d^2}{T} > \frac{2}{T}.
\end{cases}
\]

Whereas Strategy 1 outperforms Strategies 2 and 3 when

\[
\begin{cases} 
\phi < \theta \\
\frac{cv_d^2}{T} < \frac{2}{T}.
\end{cases}
\]

Proof.

First, is easy to see \(\Delta_{s_1} < \Delta_{s_2} \leq 0\) holds when

\[
\frac{2\gamma_i(\mu_i(1 - \phi) - 1 + E[\phi^2])}{(1 - \phi)^2} > 0
\]

Substituting (4) into the above inequality, reduces it to

\[
\phi > \theta
\]

When (47) holds, the difference between the estimate of Strategy 2 and Strategy 3 and the lead-time demand variance is shorter than the difference between the lead-time demand variance and Strategy 1’s estimate. All three strategies underestimate the lead-time demand variance but Strategy 2 and Strategy 3 perform better and therefore are recommended.

Second, solving \(\Delta_{s_2} < \Delta_{s_1} \leq 0\), we get

\[
\begin{cases} 
\frac{cv_d^2}{T} \leq T \\
\phi < \theta
\end{cases}
\]

This indicates that Strategy 1 produces more accurate estimates than Strategy 2 and 3, when all three strategies underestimate the lead-time demand variance.

Third, solving \(0 < \Delta_{s_1} < -\Delta_{s_2}\), we have

\[
\begin{cases} 
T < \frac{cv_d^2}{T} < \frac{2}{T} \\
\phi < \theta
\end{cases}
\]

This means that while Strategy 1 overestimates the lead-time demand variance and Strategies 2 and 3 underestimate it, the estimate of Strategy 1 is more accurate and thus is recommended over the other strategies in this case.
Fourth, solving $0 < -\Delta_{S2} < \Delta_{S1}$, we get
\[
\begin{cases}
  cv_d^2 > \frac{2}{T} \\
  \theta < \theta
\end{cases}
\]  

This suggests that Strategy 1 overestimates the lead-time demand variance while Strategy 2 and Strategy 3 underestimate it. However, Strategy 2 and Strategy 3 produce more accurate estimates than Strategy 1.

□

5. Numerical investigation

In this section, we first conduct a numerical investigation to compare the performance of the three strategies when estimating the variance of lead-time demand using SES forecasting method. Next, we compare their inventory control performance.

5.1. Variance estimates comparison

For the purpose of comparing variance estimates, we consider the actual variance of the lead-time demand given by (12) as the benchmark and analyse the (percentage) variance difference when using the variance as estimated by Strategy 1, Strategy 2 or Strategy 3. The percentage difference achieved when Strategy 1, 2 and 3 are used is denoted by $\Delta_{S1/LTD}$, $\Delta_{S2/LTD}$, and $\Delta_{S3/LTD}$, respectively and expressed as follows:

\[
\Delta_{S1/LTD} = \frac{\text{var}(LTFE)_{S1} - \text{var}(D^L)}{\text{var}(D^L)}
\]  

\[
\Delta_{S2/LTD} = \frac{\text{var}(LTFE)_{S2} - \text{var}(D^L)}{\text{var}(D^L)}
\]  

\[
\Delta_{S3/LTD} = \frac{\text{var}(LTFE)_{S3} - \text{var}(D^L)}{\text{var}(D^L)}
\]

For the purpose of testing the impact of the type of the lead-time distribution on the performance of the strategies, two distributions are considered, namely: the Geometric with parameter $p$ (including the first success), denoted by $G(p)$, and the discrete Uniform with parameters $a$ and $b$, denoted by $U[a,b]$ (Teunter et al., 2010; Syntetos et al., 2016b; Baker and Kharrat, 2018).

The detailed results of the numerical investigation are generated for SES with $\alpha = \beta = 0.1$. These results are presented for demonstration purposes. We also report some numerical results for the case $\alpha = \beta = 0.01$ to relate them to the analytical results of Propositions 5–7. Results for other values
of $\alpha$ and $\beta$ have been generated but are not reported here as they do not offer additional insights. Demand follows an ARMA(1, 1) process where $\phi$ and $\theta$ are between -1 and +1. We assume that the mean demand is $\mu = 100$ and $\sigma = 50$. These numerical values are the same with those considered in many numerical investigations in the supply chain forecasting literature (Lee et al., 2000, Ali et al., 2012). The settings described in this subsection facilitate the reproduction (Boylan, 2016) of our results should one wish to do so. In order to better link the comparative performance of the strategies to the autocorrelation of the demand, we show in Figure 1 the autocorrelation associated with the ARMA(1,1) process.

First, we discuss the impact of demand autocorrelation and mean lead-time. We consider the cases of $G(0.2)$ and $G(0.5)$ which imply average lead-times of 5 and 2 periods, respectively. Figure 2 shows the percentage difference $\Delta_{x1/LTD}$, $\Delta_{x2/LTD}$, and $\Delta_{x3/LTD}$ when $\phi$ is between -1 and +1 and $\theta$ takes the
values of 0.5, 0 and -0.5. Other results when $\theta$ varies between -1 and +1 are reported in Appendix A. In addition, we present in Figure 3 the superiority region (i.e. when the percentage difference of the variance estimate is closer the zero) of each strategy when $\theta$ and $\theta$ vary between -1 and +1. The superiority region of each strategy is given with a particular colour. The white colour is used when strategy 3 is the best, the gray shows the region where strategy 2 is the best and the dark gray is used when strategy 1 is the best.

Figure 2. Percentage variance difference of the lead-time forecast error when using the three strategies.
under a Geometric distribution

Figure 3. Accuracy superiority regions of the three strategies under a Geometric distribution

The results in Figure 2 show that in the majority of cases $\Delta_{S1/LTD}$, $\Delta_{S2/LTD}$, and $\Delta_{S3/LTD}$ are negative, which means that the three strategies underestimate the variance of the lead-time demand. However, there are a few cases where the three strategies may lead to overestimation, which mainly occur for low lead-times, low lead-time variability and highly negative values of $\phi$. This can be shown in Figure 2b for $p = 0.5$ when $\theta = 0.5$ and $\phi < -0.7$. Similar findings are presented in Propositions 2–3 for the i.i.d. demand case and in Propositions 5–7 for the auto-correlated demand case with $\alpha, \beta \to 0$. It should be noted, based on Figure 2, that when there is positive autocorrelation in the demand (i.e., $\phi > \theta$), Strategy 3 leads to a more accurate estimate than Strategy 1, and this is reversed for negatively autocorrelated demand (i.e., $\phi < \theta$). Similar results are analytically shown in Proposition 7 when $\alpha, \beta \to 0$.

Figure 3 shows that the biggest superiority region is associated with Strategy 2, suggesting that this is often more accurate than both Strategies 1 and 3. Consistent with Proposition 7, Figure 3 also shows that when $\phi < \theta$, Strategy 1 is often more accurate than Strategy 2. There are a few cases with high negative autocorrelation (i.e. for $\phi < \theta$ and highly negative values of $\phi$) where Strategy 3 is more accurate than Strategy 2 (as shown in Figure 3b). Note that the relatively inferior performance of Strategy 3 in cases of negative autocorrelation can be explained by the fact that demand aggregation decreases considerably the lead-time demand variance estimate as shown in Rostami-Tabar et al. (2014), which subsequently deteriorates the estimate of the lead-time demand variance.

For $\phi = \theta$, regardless of the demand autocorrelation parameters and lead-time values, Strategy 1 and
Strategy 3 lead to the same estimates (see Figure 2). In addition, the estimate of both Strategies 1 and 3 is less accurate than that of Strategy 2. Note that in the particular setting of an i.i.d. demand (when $\phi = \theta = 0$), Figures 2c and 1d enable us to confirm Propositions 1–3 since it is easy to check that under the considered numerical values, the squared coefficient of variation of the demand $cv_d^2 = 0.25 < \frac{z - a}{a} \left( \frac{\sigma^2}{\mu^2 + \sigma^2} \right) = 8.44$, for $p = 0.2$. This also means that all strategies underestimate the lead-time demand variance, and Strategy 2 leads to a better estimate than Strategies 1 and 3.

Furthermore, we report in Appendix B the numerical results for the case $\alpha = \beta = 0.01$. The results in Figure B1 show that, overall, the comparative performance of the strategies remains the same for low values of $\alpha$ and $\beta$. They also confirm the findings stated in Proposition 6. In fact, for low values of $\alpha$ and $\beta$, Strategy 2 and Strategy 3 always underestimate the lead-time demand variance whereas Strategy 1 may overestimate it.

We now consider the case of a uniform distribution with $U[1,9]$ and $U[1,3]$, i.e. an average lead-time equal to 5 and 2 periods, respectively. Figure 4 shows the percentage difference $\Delta_{S1/LTD}, \Delta_{S2/LTD}$, and $\Delta_{S3/LTD}$ when $\phi$ is between -1 and +1 and $\theta$ taking the values of 0.5, 0 and -0.5. The results when $\theta$ is between -1 and +1 are reported in the Appendix A. The superiority regions of the three strategies under the Uniform distribution are reported in Figure 5.

![Graph showing percentage difference for different strategies under Uniform distribution](image)

The graphs illustrate the percentage difference $\Delta_{S1/LTD}, \Delta_{S2/LTD}$, and $\Delta_{S3/LTD}$ when $\phi$ is between -1 and +1 and $\theta$ taking the values of 0.5, 0 and -0.5. The superiority regions of the three strategies under the Uniform distribution are reported in Figure 5.
Figure 4. Percentage variance difference of the lead-time forecast error when using the three strategies under a Uniform distribution

Figure 5. Accuracy superiority regions of the three strategies under a Uniform distribution
Figure 4 shows again that in most cases the three strategies underestimate the variance of the lead-time demand. Few exceptions occur for highly negative autocorrelated demand (i.e., highly negative values of $\phi$). However, it should be noted that under the Uniform lead-time distribution, it is more likely that the three strategies overestimate the lead-time demand variance as compared to the case of the Geometric distribution. This is expected since over-estimation occurs mainly for low average lead-times with low variability, which is more likely to occur under the Uniform distribution of lead-times.

The results under the Uniform distribution also show that when there is positive autocorrelation in the demand (i.e., for $\phi > \theta$), Strategy 3 leads to a more accurate estimate than Strategy 1. The opposite is true under negative autocorrelation. In addition, Strategy 2 is often more accurate than both Strategies 1 and 3 (i.e., Figure 4 shows that Strategy 2 is associated with the largest superiority region). Such outperformance of Strategy 2 is more pronounced for higher lead-times, but the gap between Strategy 2 and the two other strategies is lower in the case of the Uniform distribution than that of the Geometric. It is evident from Figure 4 that when $\phi = \theta$, Strategy 1 and Strategy 3 lead to the same result. Note that the comparative results of Strategy 2 and Strategy 3 confirm the findings of Rostami-Tabar et al. (2014) (based on their comparison of the aggregation and non-aggregation strategies under a constant lead-time). In fact, for a moderately positive or negative autocorrelation, Strategy 2 is often more accurate than Strategy 3. However, for very high positive autocorrelation (i.e. very high values of $\phi$) or high negative autocorrelation (i.e. very high values of $\theta$), as shown in Figure 5, Strategy 3 may lead to the best performance.

5.2 Analysis of inventory control performance

5.2.1 Performance evaluation methodology

In order to evaluate the inventory control performance of each strategy, we assume that the inventory is controlled according to an order-up-to-level (OUTL) policy. We present in this section the method we use in order to derive the inventory distribution under this policy for a stochastic lead-time. The inventory distribution is then used to calculate the expected cost and service level of each strategy. Note that this method is based on the work of Wang and Disney (2017), which is restricted to the case of a bounded discrete lead-time distribution.

At any period $t$, under the OUTL policy, the order $o_t$ made to replenish the inventory can be written as:
\[ o_t = \mu_L f_t + ss - i_t - w_t = \mu_L (f_t - f_{t-1}) + d_t \]  \hspace{1cm} (54)

where \( \mu_L \) is the average lead-time, \( o_t \) is the order quantity, \( ss \) is the safety stock, \( i_t \) is the inventory level, and \( w_t \) is the work-in-progress. From (54) we can directly have

\[ i_t = \mu_L f_t + ss - o_t - w_t \]  \hspace{1cm} (55)

The method used to characterise the distribution of \( i_t \) is presented in Appendix C. Based on (C12), we can derive the expected cost and service level of each strategy, which are used for the purpose of the inventory performance numerical investigation.

### 5.2.2 Inventory control performance results

In order to generate the numerical results of the inventory performance when SES is used, we consider the same numerical values of the demand process and forecast method presented in Section 5.1. We assume that \( \mu = 100, \sigma = 50 \) and \( \alpha = \beta = 0.1 \). In addition, we assume that the unit inventory holding cost \( h = 0.1 \) and the unit backlog cost is \( b = 0.9 \), which is equivalent to a target service level \( h/(h+b) = 90\% \). Note that the analysis is conducted only for the Uniform distribution since the evaluation methodology previously discussed holds only for a bounded discrete lead-time distribution.

In order to analyse the comparative inventory control performance of the three strategies, we first show in Figures 6 and 7 the expected total inventory cost and service level of the three strategies for some values of \( \phi \) and \( \theta \). Then, we present in Figure 8 the cost superiority region of each strategy for the entire range of \( \phi \) and \( \theta \).

![Figure 6. Total inventory cost of the three strategies (\( \theta = 0.5, h = 0.1, b = 0.9 \))](image-url)
Strategies 2 and 3 are associated with very similar variance estimates (as shown in Figures 4a and 4b) for low lead-times, which in turn leads to a very close inventory cost performance between them (as shown in Figure 6). Moreover, the cost difference between the two strategies increases with the lead-time. Further, Figure 6 shows that when $\phi$ takes moderate values (i.e., values between -0.6 and 0.5), Strategy 1 leads to the lowest total inventory cost. This is in line with the variance estimation results presented in Figure 5 where the superiority region of Strategy 1 is given by the dark grey colour. However, it is clear from Figure 6 that for highly positive or negative values of $\phi$, Strategies 2 and 3 are associated with lower costs than Strategy 1, and Strategy 2 often leads to the lowest cost. Furthermore, it is worth noting that the higher estimate of the variance of the lead-time demand given by Strategies 2 and 3 for high values of $\phi$ ($\phi > 0.5$), result in a higher service level as compared to that given by Strategy 1. However, for lower values of $\phi$, Figure 4 shows an opposite behaviour which is sustained in Figure 7.

Figure 8 shows that Strategy 2 is associated with the largest superiority region, which means that for a wider range of the demand process parameters, it leads to inventory costs that are lower than those
resulting from Strategies 1 and 3. This outperformance of Strategy 2 increases with the lead-time, further verifying the previous analysis on variance estimation. For very high positive autocorrelation (i.e. very high values of $\phi$) or high negative autocorrelation (i.e. very high values of $\theta$), Strategy 3 may lead to the lowest inventory costs. Finally, it should be noted that Strategy 1 is associated with the lowest inventory costs for negative autocorrelation (i.e. when $\phi < \theta$ but $\phi$ and $\theta$ are close). These results also confirm the findings presented in Section 5.1 where the variance estimates are analysed.

To conclude, for high positive autocorrelation, Strategy 1 is associated with the worst performance (i.e. higher costs and lower service levels) and Strategies 2-3 should be preferred.

In Figure 8, we have also incorporated some information about the empirical demand data used by Rostami-Tabar et al. (2014). The empirical dataset used for the purposes of that work came from a major European supermarket and contains demand date for 1,798 SKUs, 91 of which are identified as ARMA(1,1) processes. Figure 8 scatter-plots these 91 series (according to their auto-regressive and moving average parameter values) demonstrating the empirical relevance of Strategies 2 and 3 (which would have performed best if they were to be used for these SKUs).

6. Conclusion

In this paper, we have considered the implications of three strategies for estimating the variance of the lead-time demand (and thus calculating safety stocks), under an ARMA(1,1) demand process and a stochastic lead-time. Strategy 1 represents an aggregation of the per period variances of forecast error. Strategy 2 relates to the variance of the aggregated forecast error, and Strategy 3 considers the variance of the forecast error by temporal aggregation. We have derived analytical expressions under the three strategies and have validated previous analytical results shown in the literature in the case of i.i.d. demand. Properties of the three strategies are also discussed for certain combinations of the smoothing constant values, demand autocorrelation and lead-time variability, and a numerical investigation is conducted to compare the performance of the strategies in a broad range of settings.

Our analytical results show that for i.i.d demand with stochastic lead-times, by using the MMSE forecasting method, the three strategies lead to the same (under)estimation of the lead-time demand variance. Under SES, Strategy 1 (or equivalently Strategy 3 if the same smoothing constant value is used) and Strategy 2 always underestimate the lead-time demand variance when the demand and the forecasts are associated with low variability and the lead-time is highly variable. The conditions presented in this paper, under which one strategy outperforms the others, can help practitioners choose between Strategy 1 and Strategy 2 by simply comparing the coefficient of variation of demand and a function of the lead-time mean and variance.

Under SES, Strategy 2 is a potential champion in the real world. Overall, it leads to the most accurate estimates of the lead-time demand variance, followed by Strategy 3. Strategy 3 does leads to the most
accurate estimate and the best inventory performance (i.e. lower inventory cost and higher service level) under a high positive autocorrelated demand and in few cases where demand data exhibit high negative autocorrelation. There are some exceptions, where Strategy 1 leads to the most accurate estimate and lower inventory costs, mainly for negatively autocorrelated data. Furthermore, under the MMSE forecasting method, Strategy 1 leads to the best accuracy and inventory performance for a negative autocorrelated demand, with some exceptions where Strategy 3 is the best (especially for higher lead-times); whereas under a positive autocorrelation, Strategy 3 is always associated with superior performance.

Managers need to be aware that the classical strategy (Strategy 1) is associated with the least accuracy and inventory performance when demand is positively autocorrelated. Moreover, some analysis of empirical data shows the prevalence of positive autocorrelated demand and thus the fact that such SKUs would benefit from the implementation of Strategy 2 or Strategy 3. Note that the outperformance of Strategy 2 is more pronounced for higher lead-times. Finally, the numerical results show that the comparative performance of the three strategies is consistent regardless of the lead-time distribution.

In terms of future work, it would be interesting to empirically analyse the comparative (accuracy and inventory) performance of the three strategies. Another interesting avenue for further research would be to extend the work conducted in this paper to other ARIMA-type demand processes. It should be noted that although our results have been derived under ARMA(1,1) demand, the three strategies considered constitute general strategies that can be used to calculate the variance of lead-time demand forecast error under other ARMA processes, although the analytical derivations may, in some cases, not be tractable.

**Appendix A. Numerical results for the entire range of \( \phi \) and \( \theta \)**
Figure A1. Percentage variance difference when using the three strategies under a Geometric distribution, $G(0.2)$ ($\alpha = \beta = 0.1$)

Figure A2. Percentage variance difference when using the three strategies under a Geometric distribution, $G(0.5)$ ($\alpha = \beta = 0.1$)

Figure A3. Percentage variance difference when using the three strategies under a Uniform distribution, $U[1,9]$ ($\alpha = \beta = 0.1$)
Figure A4. Percentage variance difference when using the three strategies under a Uniform distribution, $U[1,3]$, ($\alpha = \beta = 0.01$)

Appendix B. Numerical results for $\alpha = \beta = 0.01$ (Geometric distribution)
Appendix C. Derivation of the inventory distribution

At any period $t$, the inventory level is expressed as:

$$i_t = \mu_t f_t + ss - o_t - w_t$$  \hspace{1cm} (C1)

where $\mu_t$ is the average lead-time, $o_t$ is the order quantity, $ss$ is the safety stock, and $w_t$ is the work-in-progress.

Only $w_t$ in the right hand side of equation (C1) is affected by the realisation of lead-time, while the other terms are only related to the mean and/or variance of lead-time. Therefore, it is $w_t$ that we need to discuss further.
We need to introduce the concept of outstanding status here. Suppose we are at time \( t \), and we look at the orders placed in the past \( L^+ - 1 \) periods (\( L^+ \) is the maximum lead-time and \( L^- \) is the minimum lead-time). The reason that we don't need to look further back is that all orders have been completed at time \( t \). The outstanding status simply indicates: out of all these \( L^+ - 1 \) orders, which ones are completed and which ones are not. It can be represented by a vector \( \xi \) that contains \( L^+ - 1 \) elements that are either 0 or 1. For instance, \( \xi(1) = 1 \) means that the order placed one period ago is outstanding; \( \xi(3) = 0 \) means that the order placed three periods ago is already completed. \( \xi \) has \( 2^{L^+-L^-} \) possible realisations, which means that the outstanding status has \( 2^{L^+-L^-} \) possibilities. The work-in-progress is then the sum of all previous orders that are outstanding:

\[
W_t = \sum_{\xi(k) = 1}^{X} o_{t-k}
\]  

(C2)

For the computation procedure, the idea is that the inventory distribution is a mixture of \( 2^{L^+-L^-} \) distributions, each of which possesses the same outstanding status. All the periods that have the same outstanding status have the same work-in-progress distribution. We denote the distribution components by \( \psi(\xi, 1) \) (as they can be solely identified by \( \xi \)). The steps of the computation procedure are as follows:

1. Calculate the mean, variance and covariance functions.

The means are easy to derive from (C1):

\[
E(i, \xi) = ss + (\mu_L - \xi \mathbf{1} - 1) \mu_d
\]  

(C3)

where \( \mathbf{1} \) is an all-one vector with proper dimensions. For the variance, from (C1) we have

\[
Var(i; \xi) = \mu_L \tau \var(f) + \var(o) + 2\mu_L \cov(f, o) + \var(w) - 2\mu_L \cov(f, w)
\]  

(C4)

Note that work-in-progress is the random sum of past orders. The terms with the variable of work-in-progress are affected by the random lead-time. Therefore for the last three terms in (39), we need the covariance functions w.r.t. time delay between forecast and orders \( \cov(f_t, o_{t-k}) \), and between orders, \( \cov(o_t, o_{t-k}) \). Formally, we can write an \( (L^+ - 1) \times (L^+ - 1) \) matrix \( \Gamma \) and two \( (L^+ - 1) \times 1 \) vectors \( G, H \), where their elements are:

\[
\gamma_{jk} = \cov(o_t, o_{t-j+k})
\]  

(C5)

\[
g_k = \cov(f_t, o_{t-k})
\]  

(C6)
From (C1), it is easy to see that

\[ h_k = \text{cov}(o_t, o_{t-k}) \]  \hfill (C7)

\[ \text{cov}(w) = \xi \Gamma \xi \]  \hfill (C8)

\[ \text{cov}(f, w) = \xi G \]  \hfill (C9)

\[ \text{cov}(o, w) = \xi H \]  \hfill (C10)

To summarize, the covariances needed to calculate the component distributions are \( \text{var}(f) \), \( \text{var}(w) \), \( \text{cov}(f_t, o_{t-k}) \) and \( \text{cov}(o_t, o_{t-k}) \), which are not difficult to derive from the expressions of \( d_t \), \( f_t \) and \( o_t \).

2. Calculate the probability of \( \xi \).

For each realization of outstanding status, the probability is calculated by the following equation:

\[ p(\xi) = \prod_{k=1}^{L^+ - 1} [1 - \xi(k)] \Psi_L(k) + \xi(k) \bar{\Psi}_L(k) \]  \hfill (C11)

where \( \Psi_L(k) \) and \( \bar{\Psi}_L(k) \) are the cumulative distribution function and the complementary cumulative distribution function of the discrete lead-time, respectively.

3. The component distribution \( \psi(\xi; 1) \) identified by \( \xi \) follows a normal distribution with \( \text{E}(i; \xi) \) and \( \text{var}(i; \xi) \) as mean and variance. The overall distribution is then

\[ \psi(i) = \sum_{\xi} p(\xi) \psi(i; \xi) \]  \hfill (C12)
References


Axšäter, S. 2006. Inventory Control. 2nd ed. Springer-Verlag, New York.


