Approximations for the Lead Time Variance: a Forecasting and Inventory Evaluation

Patrick Saoud[a,*], Nikolaos Kourentzes[b], John E. Boylan[a]

[a]*Lancaster University Management School*
Department of Management Science, Lancaster, LA1 4YX, United Kingdom
[b]*Skövde Artificial Intelligence Lab, School of Informatics, University of Skövde, Sweden.*

Abstract

Safety stock is necessary for firms in order to manage the uncertainty of demand. A key component in its determination is the estimation of the variance of the forecast error over lead time. Given the multitude of demand processes that lack analytical expressions of the variance of forecast error, an approximation is needed. It is common to resort to finding the one-step ahead forecast errors variance and scaling it by the lead time. However, this approximation is flawed for many processes as it overlooks the autocorrelations that arise between forecasts made at different lead times. This research addresses the issue of these correlations first by demonstrating their existence for some fundamental demand processes, and second by showing through an inventory simulation the inadequacy of the approximation. We propose to monitor the empirical variance of the lead time errors, instead of estimating the point forecast error variance and extending it over the lead time interval. The simulation findings indicate that this approach provides superior results to other approximations in terms of cycle-service level. Given its lack of assumptions and computational simplicity, it can be easily implemented in any software, making it appealing to both practitioners and academics.

Keywords: Forecasting, lead time demand variance, demand uncertainty, safety stock, forecast errors correlations

1. Introduction

With an increase in competitiveness, meeting customer demand has become a target that businesses strive to achieve. Maintaining a balance between excess inventory and lost sales is a necessity for better performance. Safety stock plays a pivotal role in this, as it allows buffering against demand uncertainty. In practice, safety stock is determined by multiplying the standard deviation of forecast error over lead time by the inverse of the distribution function that represents these errors at the desired level of coverage. The forecast error over lead time implies knowledge of the cumulative forecast over the same

*Corresponding author.
Email address: patricksaoud@gmail.com (Patrick Saoud)

period for which variance expressions are not readily available. This leads researchers and practitioners to either impose severe assumptions or use approximations to arrive at the desired variance by using the point forecast errors. Textbooks often prescribe scaling the variance for the one-step ahead forecast errors by the lead time as an approximation (Axsäter, 2015). Due to its simplicity and ease of application, this formula is employed frequently, requiring only the calculation of the one-step-ahead forecast errors variance as an input. This method suffers from a serious drawback as it fails to capture the correlations between forecast errors (Johnston and Harrison, 1986; Barrow and Kourentzes, 2016). Not accounting for these correlations leads to the variance of errors being under-estimated, which in turn results in inappropriate safety stocks levels being determined.

This research acknowledges the existence of these correlations and provides some new insights on their effect on forecasts, service levels and inventory holdings. While these correlations are the motivation for this paper, the research aims at examining the performance of three different forecasting methods for the variance of lead time forecast error on stock control levels under various types of forecasting uncertainty. Second, within an AutoRegressive Integrated Moving Average (ARIMA) framework, we examine analytically simple fundamental Normal demand processes and use the variance-covariance matrix to illustrate the validity of the traditional approximation for an ARIMA(0,0,0) model but highlight the approximation's inadequacy for other commonly applied models due to the appearance of these correlations. This occurs even when the demand model and parameters are assumed to be fully known, as they appear in the variance-covariance matrix of the forecast errors, irrespective of the assumptions on the distribution of the latter, one of which is typically independence. The form of these correlations depends on the model structure and its parameters. Third, we also recommend the use of a simple yet intuitive heuristic for approximating the forecast errors variance, which consists of monitoring the forecast errors distribution over the lead time, rather than resorting to the approximation discussed before. While this approach has appeared before in the literature, the underlying motivation behind its choice has not been discussed, and this paper seeks to justify its use and advantage over other approximations. The heuristic relies on the empirical cumulative forecast errors over lead time directly, instead of building on the one (or multiple) step ahead forecast errors.

In contrast with existing research (e.g. Prak et al., 2017), we do not seek to determine remedies for specific demand processes where the forecasting model is mis-specified; rather we aim at studying the performance of approximations for estimating the lead time variance of forecast errors. Since knowledge of the underlying process is impossible in practice, we proceed to examine three competing approximations, explaining the rationale behind them and linking them to the analytical insights we provide. This paper contributes to the existing literature by evaluating the inventory implications of these approximations, under different forecasting uncertainty settings and data generating processes, including both stationary and non-stationary ones, discussing the conditions under which each approximation is viable. We validate our analytical and simulation results on a real case study, using data from a US retailer. Our empirical findings indicate that this method achieves superior cycle-service level results compared to alternatives, which coupled with its ease of implementation, underlines its usefulness for research and practice.

The rest of this paper is organised as follows: Section 2 first discusses demand uncertainty and the approaches to model it; second, it covers the correlations between forecast errors over a lead time and how these impact the estimation of the variance of lead time forecast errors. Section 3 shows the existence of these correlations from a theoretical standpoint for simple yet fundamental demand processes. Section 4 examines the different approximations in estimating lead time forecast errors variance, while Section 5 reports the findings drawn from an inventory simulation on the adequacy of these approximations, followed by the results gained from using real data in Section 6.

2. Background Literature

2.1. Demand Uncertainty and Variability

Demand uncertainty refers to the unpredictability that arises in forecasting future demand, as opposed to variability which is defined as the fluctuations of demand around its mean. It is represented by the distribution of the forecast errors, and its impact on safety stocks is quantified by the variance of the latter. Failure to acknowledge this difference results in setting inappropriate safety stocks. The forecast uncertainty can be split into three types, which are reviewed subsequently. We show in the next section that these uncertainties appear in the calculation of the variance of lead time demand forecast errors.

In modeling demand, we are confronted with three types of uncertainty, endemic to any forecasting problem: *model, parameter* and *sample size* (Chatfield, 1995). Any forecasting task faces model uncertainty, as in reality the underlying Data Generating Process (DGP) is unknown and it is impossible to diagnose how closely this is approximated by a specified model. Many forecasters and inventory researchers overlook this and fail to account for this uncertainty, which is reflected in an inadequate estimation of the variance of forecast error. As a result, inappropriate safety stocks are set and higher inventory costs are incurred (Badinelli, 1990; Kim and Ryan, 2003; Dong and Lee, 2003). Even if the form of the true DGP is assumed to be known, it is questionable whether the parameters can be perfectly known. Parameter Uncertainty refers to this, where the parameters are misspecified, which can affect the performance of demand prediction intervals (Lee and Scholtes, 2014) as well as safety stocks (Ritchken and Sankar, 1984). Sample Size Uncertainty refers to the case where both model and parameters are known, and the uncertainty is from sampling issues, which results in a disparity between the asymptotic properties of the model and the finite sample properties of the underlying data. In practice, it is often difficult to distinguish between the three types of uncertainty. For example, a misspecified parameter can make a model term insignificant, and change the specified model as well.

In a stock control context, quantiles are required in addition to point forecasts in order to model forecast uncertainty. Uncertainties, which generate biases, are not always unfavourable, as a parameter being optimal in terms of Mean Squared Error (MSE) is not optimal in an inventory context (Janssen et al., 2009), due to the difference in objective functions (Strijbosch et al., 2011). For safety stock purposes, the variance of the forecast errors is required as an input, despite many different procedures being proposed in the literature (a comprehensive review can be found in Gonçalves et al., 2020). Since the error distribution is part of the forecasting model, the uncertainty surrounding the size of its standard deviation should be factored in, as it is estimated rather than determined a priori. Thus, all three uncertainties include this estimation uncertainty. While the impact of these uncertainties has featured separately in stock control papers, they have not been compared in terms of inventory performance. In this research, these uncertainties will be considered, in order to assess the effect of the safety stock approximations under these different uncertainty scenarios.

The uncertainty surrounding demand is quantified by constructing demand prediction intervals to confine the possible regions between which future demand might lie. The first step consists of determining the variance of the forecast errors. However, since many of the tasks consist of finite and limited data, the conditional variance of the i-th step ahead forecast error is employed ($\sigma^2_{t+i|t}$), as opposed to the unconditional one (σ^2_{t+i}) which provides the asymptotic value. Henceforth, the variance calculated is conditional on the data made available up until the estimation time t. This is the theoretical variance; nevertheless, when forecasting, the variance component of the forecast errors needs to be calculated, and thus $\widehat{\sigma}^2_{t+1|t}$ replaces $\sigma^2_{t+1|t}$ for estimation purposes. After determining the variance component, the intervals are built with the use of the corresponding percentile of the assumed error distribution. The construction of demand or prediction intervals falls into three categories. The parametric stream of the research assumes that the underlying DGP can be modeled by a forecasting model (Lee, 2014), which depends on the researcher having adequately approximated the true model. For example, the Exponential Smoothing family of models can be used to manage SKUs with different patterns (Snyder et al., 2002), while knowing that it might not be optimal for all (if any) the time series; nevertheless its lead time expressions are derived for inventory purposes (Snyder et al., 2004). The non-parametric stream refrains from imposing any assumption on the demand process; and instead exploits the observed properties of the forecasting error density function. Examples consist of Chebyshev's Inequality (Gardner, 1988) or bootstrapping methods, with the latter finding successful use in approximating the intervals of a known demand process (Thombs and Schucany, 1990), and also performing well in setting reorder points (Wang and Rao, 1992) and meeting service levels (Fricker and Goodhart, 2000). A bootstrapping procedure for estimating safety stocks under skewed or multimodal distributions is proposed by Saldanha et al. (2020) as a remedy for demand processes that exhibit these departures from the typical assumptions surrounding demand. Similarly, Kernel Density Estimation has been used to directly estimate quantiles from the forecast error distribution to inform the calculation of safety stock (Trapero et al., 2018). The semi-parametric stream uses a mixture of the first two to model the intervals, such as those found in Taylor and Bunn (1999) and Lee and Scholtes (2014). While demand prediction intervals have been studied in a forecasting context, they bear similarities to safety stocks, with the stark difference between them being that the latter requires an estimate of the variance of forecast errors over lead time.

2.2. Safety Stock Estimation

The estimation of safety stocks requires the variance of forecast errors as input, as typically captured by the Mean Squared Error (MSE). Underlying this are two implicit assumptions. First, the forecasts are unbiased or have very little bias, which allows the variance of the errors to be approximated by the MSE (Wagner, 2002). This problem has been addressed by Lee and Scholtes (2014) for estimating better prediction intervals, and featured at the heart of the studies by Manary and Willems (2008) and Manary et al. (2009), who examined its impact on reorder points for Intel's inventories. The other assumption is that the errors are homoscedastic, which has been challenged as industrial data has exhibited some evidence of heteroscedasticity (Zhang, 2007; Stößlein et al., 2014; Trapero et al., 2019a,b).

There is another problem in estimating the forecast error variance over lead time. Indeed, forecasts are produced for several consecutive horizons in the future, akin to an overlapping temporal aggregation of demand itself over a window equal to the lead time (Boylan and Babai, 2016). This implies that the variance of the errors should cover the lead time window as well, and this brings forth a new issue which is the correlation of errors over lead times or horizons (up until this point, the terms "lead times" and "forecasting horizons" have been employed interchangeably. In this paper, the term "horizon" is used to denote the protection interval for the safety stocks, and the exact relationship between these two in a periodic review context is Horizon = Lead Time + Review Period, while in a continuous review context, Horizon = Lead Time).

In the literature there are lead time forecast errors variance expressions for the well-known models; however with the multitude of possible demand processes, the majority of cases remain uncharacterised, and therefore a heuristic is required to circumvent this limitation. Standard textbooks recommend calculating the one-step-ahead forecast errors variance and multiplying it by the lead time, or in its more familiar form: $L\widehat{\sigma}^2_{t+1|t}$. This approximation depends only on the estimated variance of the one-step ahead forecast errors, which under perfect information of the demand pattern, reduces to the model innovations. No other parameters are required, such as the demand DGP autoregressive or moving average parameters. At first sight, its simplicity and lack of assumption of DGP makes its use appealing. Chatfield and Koehler (1991) criticised this method, with their main line of attack being that the concept of lead time forecasts is often confounded with that of forecasting for a horizon. Chatfield (1993) warns against the use of this approximation, stating that it possesses no theoretical basis and does not accommodate the different properties of the prediction intervals. Koehler (1990) challenges the robustness of this method, pointing out that this equation holds in the presence of the Random Walk model for multiple-steps-ahead point forecast errors. This approximation however is the cumulative L-steps-ahead conditional variance for an independent and identical demand (i.i.d) process, and it hinges on the crucial assumption that the demand being studied can be approximated in that fashion, thus ignoring other time series patterns such as autocorrelation, trend, moving averages or seasonal patterns. Furthermore, it is questionable whether the estimated MSE in-sample errors approximates well the out-of-sample variance of the forecast errors, an important distinction that is lost if the conditionality of errors is not considered (Barrow and Kourentzes, 2016).

An important distinction must be made between model errors (i.e. the process underlying the data) and forecast method errors, as these two terms are sometimes confused. The model errors or innovations at time t, ε_t are those found in the DGP; they form the stochastic component of the demand process. The forecasting or method residuals or errors at time t, e_t, measure the difference between actuals and forecasts. Under perfect knowledge of the demand, the forecast method errors are a function of the model structure, its errors and the set of parameters Θ, i.e. $e_{t+i|t} = f(\varepsilon_{t+1}, ...\varepsilon_{t+i}, \Theta)$, and this causes the correlations between multiple-steps-ahead errors to appear for several demand models (this is elaborated further in Section 3). Indeed, the main caveat of the current procedure for estimating variance is that it fails to capture two sources of correlations: (i) correlations between errors at different forecasting horizons and (ii) correlations between forecasts over cumulative lead time (Barrow and Kourentzes, 2016). The first refers to the correlations due to mis-specifying the forecasting model as i.i.d, while the second refers to those that arise due to the forecast errors at different lead times being correlated with each other. These correlations are the emphasis of the discussion in this paper, as they are often overlooked in many of the estimation procedures, which ensues in inaccurate safety stocks levels.

For the first source of correlation, an implicit assumption being made is that the multiple-steps-ahead forecast errors all possess equal variance, with $\sigma^2_{t+i|t} = \sigma^2_{t+1|t}$. This assertion holds for the i.i.d process, $y_t = \mu + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$ and $\mathrm{Cov}(\varepsilon_{t+i}, \varepsilon_{t+j}) = 0$ for $i \neq j$. For other processes, $\sigma^2_{t+i|t} \neq \sigma^2_{t+1|t}$, as can be seen for the multiple-steps-ahead error equation for many DGPs. This already renders the approximation inappropriate. Kourentzes (2013) suggests a simple remedy to overcome this issue; namely to sum up the variance of forecast errors at different steps-ahead. Indeed, this approach consists of first estimating the conditional variance of the forecasting errors at each until the desired lead time L, and then adding them up to determine the variance at L, to give $\sum_{i=1}^{L} \sigma^2_{t+i|t}$. This approach is independent of any model assumptions, as it retrieves all the multiple-steps-ahead errors until the lead time and adds them up, and hence tackles the first issue posed by the $L\widehat{\sigma}^2_{t+1|t}$ approximation. This however does not address the correlation of forecasts at different lead times, as summing up the individual components implies automatically that the covariance terms between errors is 0.

The second source of correlation was acknowledged by Box et al. (2015) who proved its existence for ARIMA processes. The confusion between model innovations and forecast errors has partly led to these correlations being overlooked in the literature. Since the forecasts are produced for several steps-ahead and then summed, they are likely to be correlated as they share some of the information from previous periods, irrespective of whether the correct model has been identified or not. Johnston and Harrison (1986) used a Dynamic Linear Model formulation to show that ignoring these correlations could lead to an understatement of the variance of lead-time demand. They noted that this correlation existed, and they highlighted its omission from the ordinary methods of estimating the cumulative variance, attributing this correlation partly due to the need to estimate the level of the series, as well as the parameters. Prak et al. (2017) showed its existence under i.i.d demand, and provided correction terms for fitting both Simple Moving Average and Simple Exponential Smoothing procedures for this particular process, which delivered better safety stocks and service levels. While their correction terms focused on the i.i.d case and cannot be extended to other demand processes, their work showed the impact of fitting a wrong model to a specific demand (i.e. the presence of correlations between errors due to model uncertainty). Nonetheless, even under perfect knowledge of the demand, and for different processes, the correlations may still exist, due to the model structure and parameters, and this is detailed further in Section 3. More recent contributions have explored additional demand models (Prak and Teunter, 2019; Prak et al., 2021, with a focus on estimation issues) expanding the cases that we can resolve. In Section 4 we discuss general approximations of the lead time variance, complimenting these.

3. Theoretical Derivations of the Variance and Covariance Terms

In this section, we examine the conditional variance of the lead time forecast errors analytically for certain fundamental demand processes and compare it to the typical $L\widehat{\sigma}^2_{t+1|t}$ approximation, with the objective of demonstrating the inadequacy of the latter and quantifying any ensuing losses from its use. This entails extracting the variance over lead time of these errors, as well as the covariance term between the errors, thus providing a unifying framework for the expressions of the demand processes studied. The covariance terms can then be used to form the Error Variance-Covariance matrix ($\mathbf{\Sigma}$), which can display all these terms as well as demonstrate the impact of the accrual of these correlations and highlight how existing approximations capture or overlook these. We focus on DGPs that stem from the ARIMA(p,d,q) family of linear time series model, propounded by (Box et al., 2015). Its general expression is $\nabla^d y_t = \varepsilon_t \Theta(B)/\Phi(B) + c$, where y_t and ε_t represent the demand and the model innovations at time t, B the backshift operator, ∇, the difference operator defined as $\nabla = (1 - B)$, c the level or constant term, and $\Theta(B)$ and $\Phi(B)$ connote respectively the moving average and autoregressive operators in their polynomial form. The innovations ε_t adhere to an independent white noise process $N(0, \sigma^2)$. The ARIMA family encompasses many demand models studied in the literature and is common in the context of inventory control (Aviv, 2003), supporting our selection for this analysis.

While many textbooks cover the variance of point forecast errors for these processes, the cumulative aspect has received much less treatment. Nevertheless, while some of the cumulative variance expressions can be found in the literature (e.g., Ray, 1982), the core of the discussion here revolves around the covariance terms and the resulting Error Variance-Covariance matrix. We restrict our attention to basic ARIMA models, where the parameter order is fairly low, enabling relatively neat derivation of the expressions. Nonetheless, if the results hold for the basic ARIMA models, then they can be extended to more complex processes from the same family. Consequently, the following models are studied: ARIMA(0,0,0), ARIMA(0,1,0), ARIMA(0,0,1), ARIMA(0,1,1) and ARIMA(1,0,0). The inclusion of the first process, also known as i.i.d demand, will help explain the origin of the standard safety stock approximation and thus point to the inadequacy of its use for other processes differing from it. With the exception of ARIMA(0,1,1), these demand processes are the most basic i.i.d, Random Walk, Moving Average and Autoregressive processes that can be chosen; while ARIMA(0,1,1) underpins the Simple Exponential Smoothing process.

For each process, it is presumed that the correct model and parameters are known a priori, implying that all forecasts are unbiased. We show that the correlation terms are present in these cases with no model mis-specification. If the results hold under this premise, then we postulate that they hold more generally. All forecasts are conditional on the information available at t. For ease of notation, y_t will represent demand at time t, and $Y_{t+L} = \sum_{i=1}^{L} y_{t+i}$, where L is the lead time. The same logic applies for the conditional forecasts, $\widehat{Y}_{t+L|t} = \sum_{i=1}^{L} \widehat{y}_{t+i|t}$ and the forecast errors are denoted by $e_{t+i|t} = y_t - \widehat{y}_{t+i|t}$. It should be noted that since the assumption of perfect demand knowledge is imposed here, then all forecasts are unbiased, and thus $\widehat{Y}_{t+L|t} = E[\widehat{Y}_{t+L|t}]$. The cumulative errors, $E_{t+L|t}$, which is the sum of errors up to lead time L, can be expressed as $E_{t+L|t} = Y_{t+L} - \widehat{Y}_{t+L|t} = \sum_{i=1}^{L} e_{t+i|t}$. The conditional variance of the cumulative errors may be written as

$$\text{Var}\left(E_{t+L|t}\right) = \sum_{i=1}^{L} \text{Var}\left(e_{t+i|t}\right) + 2 \sum_{i=1,i<j}^{L} \text{Cov}\left(e_{t+i|t}, e_{t+j|t}\right). \tag{1}$$

If the errors are independent, then $\text{Cov}\left(e_{t+i|t}, e_{t+j|t}\right) = 0$ for $i \neq j$, so the variance of the cumulative errors would just reduce to the sum of the individual variances and $\text{Var}\left(E_{t+L|t}\right) = \sum_{i=1}^{L} \text{Var}\left(e_{t+i|t}\right)$. However, despite the disturbances ε_t being independent, the forecasting errors e_t may display a correlation over time (Johnston and Harrison, 1986; Barrow and Kourentzes, 2016; Box et al., 2015), and this is demonstrated later in this section.

$\sum_{i=1,i<j}^{L} 2\text{Cov}(e_{t+i|t}, e_{t+j|t})$, denotes the sum of all the covariances between the errors. This term differs for each process, depending on the DGP and on the forecast error structure. The conditional variance expressions for the forecast error in their variance-covariance matrix will be provided, which takes the form:

$$\underset{(L\times L)}{\mathbf{\Sigma}} = \begin{bmatrix} \text{Var}(e_{t+1|t}) & \text{Cov}(e_{t+1|t}, e_{t+2|t}) & \text{Cov}(e_{t+1|t}, e_{t+3|t}) & \ldots & \text{Cov}(e_{t+1|t}, e_{t+L|t}) \\ \text{Cov}(e_{t+2|t}, e_{t+1|t}) & \text{Var}(e_{t+2|t}) & \text{Cov}(e_{t+2|t}, e_{t+3|t}) & \ldots & \text{Cov}(e_{t+2|t}, e_{t+L|t}) \\ \text{Cov}(e_{t+3|t}, e_{t+1|t}) & \text{Cov}(e_{t+2|t}, e_{t+3|t}) & \text{Var}(e_{t+3|t}) & \ldots & \text{Cov}(e_{t+3|t}, e_{t+L|t}) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \text{Cov}(e_{t+L|t}, e_{t+1|t}) & \text{Cov}(e_{t+L|t+2|t}) & \text{Cov}(e_{t+L|t+3|t}) & \ldots & \text{Var}(e_{t+L|t}) \end{bmatrix}$$

The Variance-Covariance Matrix $\mathbf{\Sigma}$ encompasses all the information required for the computation of the conditional variance of the errors for safety stocks. For example, the $L\sigma^2_{t+1|t}$ approach consists of using one element from the matrix, the top entry which corresponds to the one-step-ahead conditional variance, and multiplying it by the lead time. Given that the matrix is square, the sum of variances estimation method is its trace, $\text{tr}(\mathbf{\Sigma})$. The true conditional variance of the cumulative errors at lead time L is the entry-wise sum of the Variance-Covariance matrix terms.

3.1. ARIMA(0,0,0)

The i.i.d demand model is a simple process and quite commonly assumed in the supply chain literature. This model is the ARIMA(0,0,0) with constant: $y_t = \mu + \varepsilon_t$. For a lead time of L, its $t + L$ actual value, forecast and forecast error expressions are $y_{t+L|t} = \mu + \varepsilon_{t+L}$, $\widehat{y}_{t+L|t} = \mu$ and $e_{t+L|t} = \varepsilon_{t+L}$, which gives $\text{Var}(e_{t+L|t}) = \sigma^2$. The term σ^2 refers to variance of the innovations. Usually, the variance of the forecast errors is not available before hand, and it has to be estimated conditionally on the data used. This estimated variance, $\widehat{\sigma}^2_{t+1|t}$, is different from the theoretical one, σ^2, and serves as a proxy for it. Nevertheless, in this section, since the focus is on theoretical derivations, the estimated variance is not needed and the variance of the innovations is used instead. The above equations for the ARIMA(0,0,0) have no parameters apart from σ, nor do they include a recursive component on y_t, and as a result, no covariance terms appear. The forecast error term $e_{t+L|t}$ is always equated with the model innovation term ε_{t+L}, and so for this case the distribution of the forecast errors is equal to the distribution of the disturbances. For cumulative demand, $Y_{t+L|t} = L\mu + \sum_{i=1}^{L} e_{t+i|t}$ and $\widehat{Y}_{t+L} = L\mu$, and $E_{t+L|t} = \sum_{i=1}^{L} e_{t+i|t}$ and $\text{Var}(E_{t+L|t}) = L\sigma^2$. For any two integers $J < K$, the covariance between two errors is $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = \text{Cov}(\varepsilon_{t+J}, \varepsilon_{t+K}) = 0$. The Variance-Covariance Matrix is

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & 0 & 0 & ... & 0 \\ 0 & \sigma^2 & 0 & ... & 0 \\ 0 & 0 & \sigma^2 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & \sigma^2 \end{bmatrix}$$

Given that $e_{t+i|t} = \varepsilon_{t+i}$ (assuming known parameters), the $L\sigma^2_{t+1|t}$ approximation is valid here. However, as we demonstrate next, the conditional variance of the cumulative errors adopts a different form for other DGPs, rendering the approximation inappropriate.

3.2. ARIMA(0,1,0)

The next model to be studied consists of the Random Walk model, ARIMA(0,1,0), or I(1); its demand is given by $y_t = y_{t-1} + \varepsilon_t$. A recursive term exists, inducing the appearance of the covariance terms, and a difference from the results pertaining to the i.i.d model. The L-steps-ahead actual values are given by $y_{t+L} = y_t + \sum_{i=1}^{L} \varepsilon_{t+i}$, and the forecast $\hat{y}_{t+L|t} = y_t$. The corresponding error is $e_{t+L|t} = \sum_{i=1}^{L} \varepsilon_{t+i}$, and its L-steps-ahead variance of the errors is $\text{Var}(e_{t+L|t}) = L\sigma^2$. For a specified lead time L, the cumulative actuals, forecasts and errors are given as $Y_{t+L|t} = \sum_{i=1}^{L} y_{t+i} = Ly_t + \sum_{i=1}^{L}(L-i+1)\varepsilon_{t+i}$, and $\widehat{Y}_{t+L|t} = E_{t+L|t} = \sum_{i=1}^{L}(L-i+1)\varepsilon_{t+i}$. Their conditional variance of the cumulative errors is:

$$\text{Var}(E_{t+L|t}) = \text{Var}\left(\sum_{i=1}^{L}(L-i+1)\varepsilon_{t+i}\right) = \left[\sum_{i=1}^{L} i^2\right]\sigma^2 = \sigma^2\left[\frac{L(L+1)(2L+1)}{6}\right]. \quad (2)$$

For any integers $J < K$, the covariance term is $\text{Cov}(e_{t+J|t}, e_{t+K|t}) = J\sigma^2$ (proof in Appendix A.1). Thus, the covariance of two terms is equal to the variance of the error with the smallest time index. The Variance-Covariance matrix for the error terms can be written as:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma^2 & \sigma^2 & ... & \sigma^2 \\ \sigma^2 & 2\sigma^2 & 2\sigma^2 & ... & 2\sigma^2 \\ \sigma^2 & 2\sigma^2 & 3\sigma^2 & ... & 3\sigma^2 \\ ... & ... & ... & ... & ... \\ \sigma^2 & 2\sigma^2 & 3\sigma^2 & ... & L\sigma^2 \end{bmatrix}$$

The Variance-Covariance matrix reveals that no entry is zero, and that they are all integer multiples of σ^2 for this process.

3.3. ARIMA(0,0,1)

The ARIMA(0,0,1) or MA(1) is written as $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$. It has short memory, retaining information from the previous shock only. To ensure the invertibility condition, we impose $|\theta| < 1$. Its L-steps-ahead actual value and forecasts are $y_{t+L} = \varepsilon_{t+L} + \theta\varepsilon_{t+L-1}$, and $\hat{y}_{t+L|t} = 0$. Its errors are $e_{t+L|t} = \varepsilon_{t+L} + \theta\varepsilon_{t+L-1}$, and its variance is $Var(e_{t+L|t}) = (1+\theta^2)\sigma^2$. For the L-steps-ahead cumulative demand, $Y_{t+L} = \theta\varepsilon_t + \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}$, and the forecast is $\widehat{Y}_{t+L|t} = \theta\varepsilon_t$. The error term is $E_{t+L|t} = \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}$, and its variance $\mathrm{Var}(E_{t+L|t}) = \left[(L-1)(1+\theta)^2 + 1\right]\sigma^2$. If $J < K$, then: $\mathrm{Cov}(e_{t+J|t}, e_{t+K|t}) = \theta\sigma^2$ if $K = J + 1$ and 0 otherwise (see Appendix A.2). For this demand, only the consecutive errors are correlated, while the non-consecutive ones are uncorrelated. The Variance-Covariance Matrix is:

$$\mathbf{\Sigma} = \begin{bmatrix} (1+\theta^2)\sigma^2 & \theta\sigma^2 & 0 & ... & 0 \\ \theta\sigma^2 & (1+\theta^2)\sigma^2 & \theta\sigma^2 & ... & 0 \\ 0 & \theta\sigma^2 & (1+\theta^2)\sigma^2 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & (1+\theta^2)\sigma^2 \end{bmatrix}$$

In this case, the covariance terms can be negative and so can their sum, depending on the sign of θ.

3.4. ARIMA(0,1,1)

The ARIMA(0,1,1) is the well-known model underpinning Simple Exponential Smoothing, which is commonly applied in practical inventory management (e.g. Graves, 1999; Babai et al., 2013; Snyder et al., 2002, 2004). It is represented by $y_t = y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. Similarly to the MA(1) process, the invertibility condition is imposed. The L-step-ahead actuals and forecasts are:

$$y_{t+L} = y_t + \theta\varepsilon_t + \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}, \tag{3}$$

and $\widehat{y}_{t+L|t} = y_t$. The L-steps-ahead forecast error is

$$e_{t+L|t} = \theta\varepsilon_t + \sum_{i=1}^{L-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+L}, \tag{4}$$

and its conditional variance of is $\mathrm{Var}(e_{t+L|t}) = [1 + (L-1)(1+\theta)^2]\sigma^2$. The L-step-ahead cumulative error is:

$$E_{t+L|t} = \varepsilon_{t+L} + \sum_{i=1}^{L-1}\left[1 + (L-i+1)(1+\theta)\right]\varepsilon_{t+i} = \sum_{i=0}^{L-1}\left[1 + i(1+\theta)\right]\varepsilon_{t+i+1} \tag{5}$$

with a conditional variance of

$$
\begin{aligned}
\operatorname{Var}(E_{t+L|t}) &= \operatorname{Var}\left(\sum_{i=0}^{L-1}[1+i(1+\theta)]\varepsilon_{t+i+1}\right) \\
&= \sigma^2 L\left[1+(1+\theta)(L-1)(\frac{(2L-1)(1+\theta)}{6}+1)\right]
\end{aligned} \tag{6}
$$

(proof in Graves (1999) or Babai et al. (2013)). For $J < K$, the covariance between two errors at different lead time is $\operatorname{Cov}(e_{t+J|t}, e_{t+K|t}) = \sigma^2(1+\theta)[1+(J-1)(1+\theta)]$ (proof in Appendix A.3). Only the J term appears in the equation, while K is absent. The Variance-Covariance matrix can be written as

$$
\mathbf{\Sigma} = \begin{bmatrix}
\sigma^2 & \sigma^2(1+\theta) & \sigma^2(1+\theta) & ... & \sigma^2(1+\theta) \\
\sigma^2(1+\theta) & [1+(1+\theta)^2]\sigma^2 & \sigma^2(1+\theta)[1+(1+\theta)] & ... & \sigma^2(1+\theta)[1+(1+\theta)] \\
\sigma^2(1+\theta) & \sigma^2(1+\theta)[1+(1+\theta)] & [1+2(1+\theta)^2]\sigma^2 & ... & \sigma^2(1+\theta)[1+2(1+\theta)] \\
... & ... & ... & ... & ... \\
\sigma^2(1+\theta) & \sigma^2(1+\theta)[1+(1+\theta)] & \sigma^2(1+\theta)[1+2(1+\theta)] & ... & [1+(L-1)(1+\theta)^2]\sigma^2
\end{bmatrix}
$$

3.5. ARIMA(1,0,0)

The ARIMA(1,0,0) or AR(1) model is a well-known demand process. Due to its simplicity, this model features in many inventory and supply chain research papers (e.g., Lee et al., 1997; Chen et al., 2000; Urban, 2005; Kahn, 1987). Its equation is $y_t = \phi y_{t-1} + \varepsilon_t$. The stationarity condition is imposed on the demand process to ~~prevent such case~~, with $|\phi| < 1$. The $t+L$ expressions for the actuals and conditional forecasts are $y_{t+L} = \phi^h y_t + \sum_{i=1}^{L} \phi^{(L-i)}\varepsilon_{t+i}$, and $\widehat{y}_{t+L|t} = \phi^h y_t$. The L-step-ahead error is $e_{t+L|t} = \sum_{i=1}^{L}\phi^{L-i}\varepsilon_{t+i}$, and its variance $\operatorname{Var}(e_{t+L|t}) = \sigma^2\frac{(1-\phi^{2L})}{1-\phi^2}$. The L-steps-ahead cumulative actuals and forecasts are $Y_{t+L} = \sum_{i=1}^{L} y_{t+i} = \sum_{i=1}^{L}\sum_{j=1}^{L-i}\left[\phi^h y_t + \phi^{(L-j)}\varepsilon_{t+i}\right]$, and $\widehat{Y}_{t+L|t} = \phi^h y_t$. The error over lead time is $E_{t+L|t} = \sum_{i=1}^{L-1}\sum_{j=1}^{L-i}\left[(1+\phi^{(L-j)})\varepsilon_{t+i} + \varepsilon_{t+L}\right]$, with a variance of

$$
\begin{aligned}
\operatorname{Var}(E_{t+L|t}) &= \operatorname{Var}\left(\sum_{i=1}^{L-1}\sum_{j=1}^{L-i}(1+\phi^{(L-j)})\varepsilon_{t+i} + \varepsilon_{t+L}\right) \\
&= \frac{\sigma^2}{(1-\phi)^2}\sum_{i=1}^{L}(1-\phi^i)^2 \\
&= \frac{\sigma^2}{(1-\phi)^2}\left[L + \frac{1-\phi^{2L}}{1-\phi^2} - 2\left(\frac{1-\phi^L}{1-\phi}\right)\right]
\end{aligned} \tag{7}
$$

For $J < K$, the covariance between these the errors at these lead times is $\operatorname{Cov}(e_{t+J|t}, e_{t+K|t}) = \sigma^2\phi^{(K-J)}\left[\frac{1-\phi^{2J}}{1-\phi^2}\right]$ (proof in Appendix A.4). Unlike the previous processes, $\operatorname{Cov}(e_{t+J|t}, e_{t+K|t})$ depends on both J and K, as opposed to solely on J, and this arises due to the inclusion of the autoregressive parameter. The Variance-Covariance Matrix can be expressed as

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \sigma^2\phi & \sigma^2\phi^2 & ... & \sigma^2\phi^{L-1} \\ \sigma^2\phi & \sigma^2\left[\frac{1-\phi^4}{1-\phi^2}\right] & \sigma^2\phi\left[\frac{1-\phi^4}{1-\phi^2}\right] & ... & \sigma^2\phi^{L-2}\left[\frac{1-\phi^4}{1-\phi^2}\right] \\ \sigma^2\phi^2 & \sigma^2\phi\left[\frac{1-\phi^4}{1-\phi^2}\right] & \sigma^2\left[\frac{1-\phi^6}{1-\phi^2}\right] & ... & \sigma^2\phi^{L-3}\left[\frac{1-\phi^6}{1-\phi^2}\right] \\ ... & ... & ... & ... & ... \\ \sigma^2\phi^{L-1} & \sigma^2\phi^{L-2}\left[\frac{1-\phi^4}{1-\phi^2}\right] & \sigma^2\phi^{L-3}\left[\frac{1-\phi^6}{1-\phi^2}\right] & ... & \sigma^2\left[\frac{1-\phi^{2L}}{1-\phi^2}\right] \end{bmatrix}$$

3.6. Discussion of Variance-Covariance Expressions

The Variance-Covariance matrix derived for the DGPs encapsulate all the elements required for estimating the true lead time variance of the forecast errors. For processes other than the ARIMA(0,0,0), the $L\sigma^2_{t+1|t}$ is an inadequate mathematical expression for the lead time variance, and this can be seen from the L-steps-ahead variance expressions being greater than it (with the exception of the ARIMA(0,0,1) with a negative θ). Indeed, inspecting the Variance-Covariance Matrix for these DGPs reveals that the parameters also appear in the lead time variance, which are omitted nevertheless in the approximation. As mentioned earlier, this approximation only utilises the first entry of the matrix to scale the variance, and this proves to be insufficient for the estimation. The sum of variance approximation is inadequate as well, as it consists of summing up the diagonal entries of the matrix, assuming that the off-diagonal entries are zero. While it makes use of more elements from the matrix and thus an improvement over the previous approximation, it nevertheless is not enough. As can be seen from the different Variance-Covariance matrices, the off-diagonal entries are not necessarily zero, and these also render this approximation inappropriate, as the pairwise covariances are discarded from the estimation under the assumption that they are equal to zero. Both approximations will generally understate the true lead time variance for the forecast errors, as they ignore many components from the matrix that enter in its calculation. The sources of variance inflation are the pairwise covariance elements, which adopt a different form based on the model structure and are usually non-zero, as well as the different variance estimates at different lead times.

4. Approximations of Lead Time Error Variance

This paper seeks to compare different approaches to approximating the variance of forecast errors over lead time. On a theoretical level the previous section showed that each demand model results in different covariance structures, and different values for the variance of errors, which depends on the demand process and its parameters. Additionally, if the parameters were estimated or the model was mis-specified, then we would expect the value of the covariance terms to be inflated as well. From a pragmatic viewpoint, the underlying DGP can assume many forms, and thus a parametric approach is not always possible. As a result, empirical approximations, independent of any assumptions of the demand process, are required. These methods approximate the lead time variance of errors from the observed forecasting errors only and do not rely on the forecasting model and its parameter. This section aims at introducing and discussing these approximations, and how they differ from each other.

4.1. Standard Approximation

As discussed in Section 2, the typical approximation is $L\widehat{\sigma}^2_{t+1|t}$, where $\widehat{\sigma}^2_{t+1|t}$ is often approximated by the in-sample MSE. This approximation has the limitations highlighted in section 2.2, but also is problematic due to the calculation of MSE, which may not represent the out-of-sample MSE, which is relevant to the inventory decisions being taken (Barrow and Kourentzes, 2016). The argument for this is that in-sample errors are naturally smaller than the forecast errors, as the forecasting method was fitted on that data and as such underestimates the observed conditional variance and ergo the future uncertainty (Makridakis and Winkler, 1989). Barrow and Kourentzes (2016) show that by switching to an out-of-sample estimate, using the MSE calculated on an appropriate validation set, the approximation becomes superior. Note that when a forecasting model is employed and estimated using Maximum Likelihood, the standard deviation of the innovation terms is an output of the estimation, which can differ from the least squares estimated value.

Alternatively, the one-step-ahead error can also be updated by a smoothing procedure (see for e.g. Syntetos and Boylan, 2006; Trapero et al., 2019a). In this work, the variance of the one-step-ahead forecast errors is calculated as:

$$\widehat{\sigma}^2_{t+1|t} = \frac{\sum_{t=1}^{n} \left(y_{t+1} - \widehat{y}_{t+1|t}\right)^2}{n} \tag{8}$$

with y_{t+1} and $\widehat{y}_{t+1|t}$ referring respectively to the realised demand and one-step-ahead conditional point forecasts.

4.2. Sum of Variances Approximation

Under this approximation, we calculate $\widehat{\sigma}^2_{t+1|t}$, $\widehat{\sigma}^2_{t+2|t}$,..., $\widehat{\sigma}^2_{t+L|t}$, with:

$$\widehat{\sigma}^2_{t+i|t} = \frac{\sum_{t=1}^{n} \left(y_{t+i} - \widehat{y}_{t+i|t}\right)^2}{(n-i+1)}, \tag{9}$$

and L is the desired lead time, and then sum them up to obtain the variance estimate $\sum_{i=1}^{L} \widehat{\sigma}^2_{t+i|t}$. This approximation falsely assumes that the errors at different steps-ahead are uncorrelated, but also suffers from a drawback due to the estimation of the multiple-steps-ahead MSE or $\widehat{\sigma}^2_{t+i|t}$. Given that multiple-steps-ahead errors are required, the forecasting model is fit on the in-sample data, and forecasts are generated using the rolling-origins procedure (Tashman, 2000). Subsequently, $\widehat{\sigma}^2_{t+i|t}$ is calculated on the multiple-step errors. Akin to the standard approximation, if only the training set is used, then the estimated variances may underestimate the future uncertainty, and out-of-sample data should be used as well. This can be achieved by using an appropriate validation set or recursively updating the estimates as new data points become available (Kourentzes, 2013). If the available sample is limited, an additional issue is that longer horizon error variance calculations may be based on very few errors, thus harming the quality of the estimation.

4.3. Cumulative Errors Approximation

While the ongoing discussion addresses the issues that are missed by the standard $L\widehat{\sigma}^2_{t+1|t}$ (and sum of variance approximation), it should be noted that it is not the norm for all researchers. For instance, Hyndman et al. (2008) provides the expression for the lead time variance of the exponential smoothing family of models, which can be used when fitting it to forecast demand in an inventory setting. Another method would consist of measuring the variance of errors over lead time, instead of calculating them at a horizon, and then aggregating them over lead time (Syntetos and Boylan, 2006; Trapero et al., 2019a). This empirical approximation is intuitive and computationally inexpensive, as it only entails calculating the cumulative forecast errors over lead time L, and then calculating its variance, or $\mathrm{Var}\left(\sum_{i=1}^{L} e_{t+i|t}\right) = \mathrm{Var}\left(\sum_{i=1}^{L}[y_{t+i} - \widehat{y}_{t+i|t}]\right)$, where y_{t+i} denotes the actual value of demand at time $t+i$, and $\widehat{y}_{t+i|t}$ its conditional forecast generated with information up until time t. Unlike the other two estimation methods, the cumulative errors approximation truly considers the variance of the lead time forecasting errors, rather than inferring for one- or multiple-steps ahead point forecasts, and this can easily be noticed when inspecting the Variance-Covariance matrices in Section 3. This approximation is not new; for instance, Eppen and Martin (1988) employ this heuristic when calculating the variance of errors, but do not provide the rationale behind its use. Lee (2014) uses this method for estimating the variance of the errors, justifying the choice as it captures the uncertainty in lead time demand, but not explaining how. More recently, Babai et al. (2021) compare the cumulative errors approach with two different lead-time variance estimation strategies involving temporal aggregation for an ARMA(1,1) demand process. Trapero et al. (2019a) also argue that using this approach offers the benefits of not needing to specify a model or its parameters. This method of tracking forecast errors seems a natural path to follow as it covers all the errors made during L; furthermore, since it contains all the information of the distribution of cumulative errors, it should mitigate the impact of the above mentioned correlations. Nevertheless, the motivation behind it has not been clearly defined in the literature. The variance-covariance matrix shown in section 3 exhibits all the components of the total variance of errors over lead time, which are featured in the cumulative error approximation. Given that the cumulative errors are comprised of the forecast errors aggregated over lead time L, we expect this approach to smooth out the errors, similarly to any other aggregation process. Finally, with the aggregation over L, we also expect the Central Limit Theorem will render the errors distributions closer to Normal, especially with higher values of the lead time (Silver et al., 2016).

The three approximations for estimating the error variance, which were elaborated in this section, will be compared via simulation in the next section, to assess their performance in an inventory context.

5. Inventory Simulation

So far, the theoretical discussion implies that even with perfect knowledge of the demand process, the errors at different lead times are correlated, and these correlations accrue in the computation of the conditional variance of the aggregate errors over lead time. Any effect on the safety stock has to account for the inventory policy. To evaluate the losses due to omitting the aforementioned covariances and the performance of the various approximations, we employ an inventory simulation.

5.1. Experimental Setup

The objective of the simulation is to compare the inventory performance of the three alternatives, outlined in section 4, namely: (i) the regular method that uses the $L\widehat{\sigma}^2_{t+1|t}$; (ii) the sum of error variances; and (iii) the cumulative error variance.

Following the theoretical discussion, we consider six DGPs, as shown in table 2. Three are stationary patterns (AR, MA and ARMA) and three are non-stationary so as to investigate both cases. While four of the above processes featured in the analytical discussion in section 3, the ARMA and ARIMA models allow the inclusion of processes with both autoregressive and moving average features of low order. The model innovations ε_t are assumed to be normally distributed, with their variance, σ, taking the values: $\{1, 5, 10, 25\}$, to represent different levels of demand volatility, and the parameters for each model are drawn from a Uniform distribution, with $0 < \phi, \theta < 1$ to impose the condition of stationarity and invertibility for each process (Box et al., 2015). We do not include the ARIMA (0,0,0) which corresponds to an i.i.d demand model, as the Variance-Covariance matrix for this process under the assumption of known DGP (i.e. SSU) in Section 3.1 shows that all three methods return the same expression for the variance of forecast errors over lead-time. Furthermore, as there are no parameters to estimate apart from the variance of the forecast errors, PU is thus equated with SSU, and the results for the effect of fitting different forecasting methods when demand is i.i.d (i.e. MU) have been studied elsewhere(see for e.g., Prak et al., 2017). For all processes we add a constant level to guarantee positive values of demand. The full set of control parameters for the experimental design are listed in the table below.

Variable	Values	Options
Variance Approximations	Regular, Cumulative, Sum	3
Forecasting Models (Uncertainty Type)	Sample Size (SSU), Parameter (PU), Model (MU)	3
Demand Process (Stationary)	AR(1), MA(1), ARMA(1,1,)	3
Demand Process (Nonstationary)	I(1), IMA(1,1), ARIMA(1,1,1)	3
Demand Noise Level	1, 5, 10, 25	4
Retailer Horizon	3, 6	2
Retailer Service Level	85%, 90%, 95%	3

Table 1: Experimental Design Control Parameters

Table 2: Simulation DGPs

	DGP	Formula
Stationary	AR(1)	$Y_t = \phi Y_{t-1} + \varepsilon_t$
	MA(1)	$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$
	ARMA(1,1)	$Y_t = \phi Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$
Non-stationary	I(1)	$Y_t = Y_{t-1} + \varepsilon_t$
	IMA(1,1)	$Y_t = Y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$
	ARIMA(1,1,1)	$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + \varepsilon_t + \theta\varepsilon_{t-1}$

For each series we generate 400 observations and split it into three subsets: the first 100 points are used as training set, where the forecasting model is fitted and the parameters are estimated; the following 200 observations are used as "burn-in", i.e. the inventory policy is allowed to run over these to eliminate the impact of any initial inventory settings (Kourentzes, 2013); and the last 100 observations constitute the test set, where the experiment outputs are recorded and results are computed. For each combination of DGP and model innovations, 500 replications were generated.

We consider the three aforementioned uncertainties: (i) Sample Size Uncertainty (SSU), where the demand model and parameters are known a priori and only σ is unknown; (ii) Parameter Uncertainty (PU) where the DGP formula is known but the parameters need to be estimated via Maximum Likelihood Estimation (MLE), and (iii) Model Uncertainty (MU), where nothing is known about the DGP and an incorrect forecasting model is deliberately fit. MU is achieved by fitting an automated ARIMA model, based on Hyndman and Khandakar (2008), and not allowing the choice of the correct DGP form for each demand type in the model identification process. Each of the three uncertainty scenarios magnifies potential problems in the estimation of the variance and therefore we can assess the impact on the inventory outcome. These are reflected in three forecasting models being fitted to each DGP, each corresponding to an uncertainty scenario.

An Order-Up-To (R,S) periodic inventory policy is considered, with a review period $R = 1$ and the measured horizons are $L + R = \{3, 6\}$. The horizons are assumed to be deterministic in order to isolate any of their variability on the total variability of safety stocks. We do not consider the one-step ahead case, as none of the discussed covariances will appear. The assumption of L>R is imposed, which is fairly typical in the literature. When an out-of-stock occurs any sales are lost; we assume that they are satisfied by a competitor. Given the difficulty in calculating the costs associated with stock-outs and lost sales, this paper will focus on setting safety stocks based on a customer service level. We track the performance at 85%, 90% and 95% target service levels, to better understand the impact of the latter on the inventory implications of each approximation. This is done by means of trade-off curves between lost sales and inventory-on-hand (Gardner, 1990), as well as realised α-cycle service level.

For each forecasting model representing an uncertainty scenario, the multiple-steps ahead point forecasts are produced following a rolling-origins scheme, and then aggregated over $L+R$ to obtain the cumulative forecasts. The safety stocks are calculated according to the estimation methods explained in section 4 for each type of approximation as:

$$SS_t = k_\alpha \widehat{\sigma}_t^M \tag{10}$$

where SS_t denotes the safety stocks at time t, k the the inverse of the normal distribution associated with the desired cycle service level α, and $\widehat{\sigma}_t^M$ the standard deviation of the forecast errors over $R+L$ for approximation M. The appropriate quantiles are constructed based on an empirical estimation of the conditional variance of the errors for all forecasting models, since the goal of this simulation is to evaluate the performance of the approximations. To determine the conditional variance of the error term, equation 9 is employed, where we use the errors in the out-of-sample subset ("burn-in" and "test" sets in our case) as it becomes available, instead of the in-sample errors, for the reasons discussed in section 4. The cumulative variance over lead time estimate, $\widehat{\sigma}_t^M$, is calculated as a rolling average over the previous forecasting errors. As the inventory simulation progresses, any new forecast errors are included in the calculation. The simulation was conducted using the R statistical language (R Core Team, 2019).

5.2. Simulation Results

5.2.1. Comparison of the three approximations

Analysing the results reveals that different σ values for the model innovations do not affect the findings greatly, and therefore we report only for $\sigma = 10$. We group the presentation of the results into the stationary and the non-stationary cases (see Table 2), as they exhibit similar behaviour. Table 3 reports the deviations of service level between realised and target, for the different horizons $L+R$, across simulation replications and DGPs. We calculate the difference from the realised service levels so that the direction of the deviations stays the same as the observed service levels, i.e. positive deviations imply larger than the target service levels. It should be noted that since service level deviations are a relative metric rather than an absolute one, we are interested in how close its magnitude is with respect to 0.

Uncertainty	Safety Stock	$\alpha = 85\%$		$\alpha = 90\%$		$\alpha = 95\%$	
		$L = 3$	$L = 6$	$L = 3$	$L = 6$	$L = 3$	$L = 6$
SSU	Regular	-2.76%	**-0.24%**	-5.18%	-3.31%	-6.89%	-5.74%
	Sum	**-1.15%**	1.36%	-3.31%	**-1.45%**	-4.87%	-3.64%
	Cumulative	2.48%	5.32%	**0.55%**	2.76%	**-0.92%**	**0.65%**
PU	Regular	-2.84%	**-0.49%**	-5.18%	-3.57%	-6.95%	-6.01%
	Sum	**-1.20%**	1.11%	-3.36%	**-1.73%**	-4.81%	-3.79%
	Cumulative	2.56%	5.35%	**0.68%**	2.93%	**-0.79%**	**0.72%**
MU	Regular	-4.31%	-1.91%	-6.94%	-5.30%	-8.77%	-8.14%
	Sum	-3.78%	**-1.16%**	-6.35%	-4.48%	-8.09%	-7.17%
	Cumulative	**2.19%**	5.25%	**0.29%**	**2.72%**	**-1.09%**	**0.56%**

Table 3: Achieved α service level deviations for stationary demand processes

Consistently across target service levels and types of uncertainty the cumulative estimation returns the highest service levels, followed by the sums and finally the regular. This is due to the underlying estimate of the error variance, which is ranked in the same order. Although this follows from the theoretical discussion, in section 3, for the stationary processes the higher service levels do not necessarily result in lower service levels deviations. At $L + R = 3$, the cumulative estimation has overall the lowest deviation from the target. Notably, as the uncertainty increases (with MU having the maximum), the deviations of the cumulative do not suffer, in contrast to the other approaches. For the regular and sum approaches, while for SSU and PU any differences in the results are minimal, the performance worsens rapidly for the MU case. Note that given the adequate estimation sample size, it is expected that there will be only small errors in the estimation of the model parameters, therefore making SSU and PU exhibit similar results. As the lead time increases ($L+R = 6$) the error variance increases as well. This is to be expected, as forecasting longer horizons is more challenging. However, when this is translated to service levels, it causes a global shift of the realised service levels upwards. This causes the cumulative approach to overshoot its target, resulting in additional stock holding than what was targeted. Noticeably, the regular approach typically under-performs in terms of meeting the cycle service level, and either sum of variances or the cumulative approaches are always preferable for achieving this end. A final useful observation is that in all cases, for the stationary time series the range of service level deviations is $[-8.77\%, 5.35\%]$, with the overshoots by the cumulative approach being quite small.

Shifting our attention to the service level deviations for the non-stationary process, provided in Table 4.The results are ordered in the same fashion. However, now in all cases the deviations are negative. The cumulative approach, for both $L + R = 3$ and $L + R = 6$, results in the minimal deviations, ranging from $[-8.24\%, -0.79\%]$ that although consistently negative, are of similar scale to the deviations observed for the stationary series. This performance is followed by the sum of variances, which again exhibits

Uncertainty	Safety Stock	$\alpha = 85\%$		$\alpha = 90\%$		$\alpha = 95\%$	
		$L = 3$	$L = 6$	$L = 3$	$L = 6$	$L = 3$	$L = 6$
SSU	Regular	-11.86%	-9.01%	-15.67%	-13.54%	-19.03%	-17.81%
	Sum of Variances	-8.38%	-5.96%	-11.55%	-9.80%	-13.88%	-13.13%
	Cumulative	**-4.32%**	**-0.84%**	**-6.71%**	**-3.80%**	**-8.24%**	**-6.09%**
PU	Regular	-11.82%	-9.01%	-15.62%	-13.51%	-18.97%	-17.77%
	Sum of Variances	-8.37%	-5.93%	-11.52%	-9.76%	-13.86%	-13.11%
	Cumulative	**-4.29%**	**-0.79%**	**-6.67%**	**-3.74%**	**-8.22%**	**-6.04%**
MU	Regular	-16.06%	-15.56%	-19.95%	-19.99%	-23.27%	-24.09%
	Sum of Variances	-14.92%	-14.14%	-18.66%	-18.19%	-21.77%	-21.89%
	Cumulative	**-6.77%**	**-4.76%**	**-8.03%**	**-6.22%**	**-7.93%**	**-6.35%**

Table 4: Achieved α service level deviations for non-stationary demand processes

large drops under full model uncertainty, that is the typical case in practice. The regular approach always performs worse, following our understanding from the theoretical discussion. Note that the deviations for the non-stationary processes, for the sum and regular approaches range between $[-24.09\%, -5.93\%]$, which is interesting to contrast with the observed range for the cumulative and the results for the stationary processes. The latter comparison demonstrates that even if in the stationary case the cumulative approach demonstrated a tendency to overshoot its target, that deviation was comparatively very small.

The service levels deviations tell only part of the full story. We refer to Figures 1 and 2 for the trade-off curves. In both stationary and non-stationary cases we observe that in terms of lost sales the regular approach returns the highest value, while the cumulative has the lowest. Furthermore, the excess inventory follows the expected outcomes, with approaches that exhibit the least lost sales naturally retaining more stock. It is interesting to observe that for a given type of uncertainty and lead time, no approach results in a trade-off curve that clearly dominates the others, only shifting the curves to a different balance point between lost sales and excess stock. Pairing this with the service levels deviations values discussed above, the better performance of the cumulative approach becomes evident. When the horizon increases from 3 to 6 periods, all trade-off curves are shifted to the right, i.e. a higher stock position. Under model uncertainty the good performance of the cumulative approach is further highlighted, as it is the only one that does not incur a substantial increase in lost sales, irrespective of the target service level.

Considering the non-stationary time series (Figure 2), we get a similar insight in the performance of the competing approaches. However, the scale of both lost sales and excess stock is increased, reflecting the difficulty in forecasting these processes. As the modeling uncertainty increases, both the regular and the sum of variances approaches deteriorate rapidly, echoing the results provided for the service levels. Comparing the curves for $L + R = 3$ and $L + R = 6$, we observe that the latter lies further on the

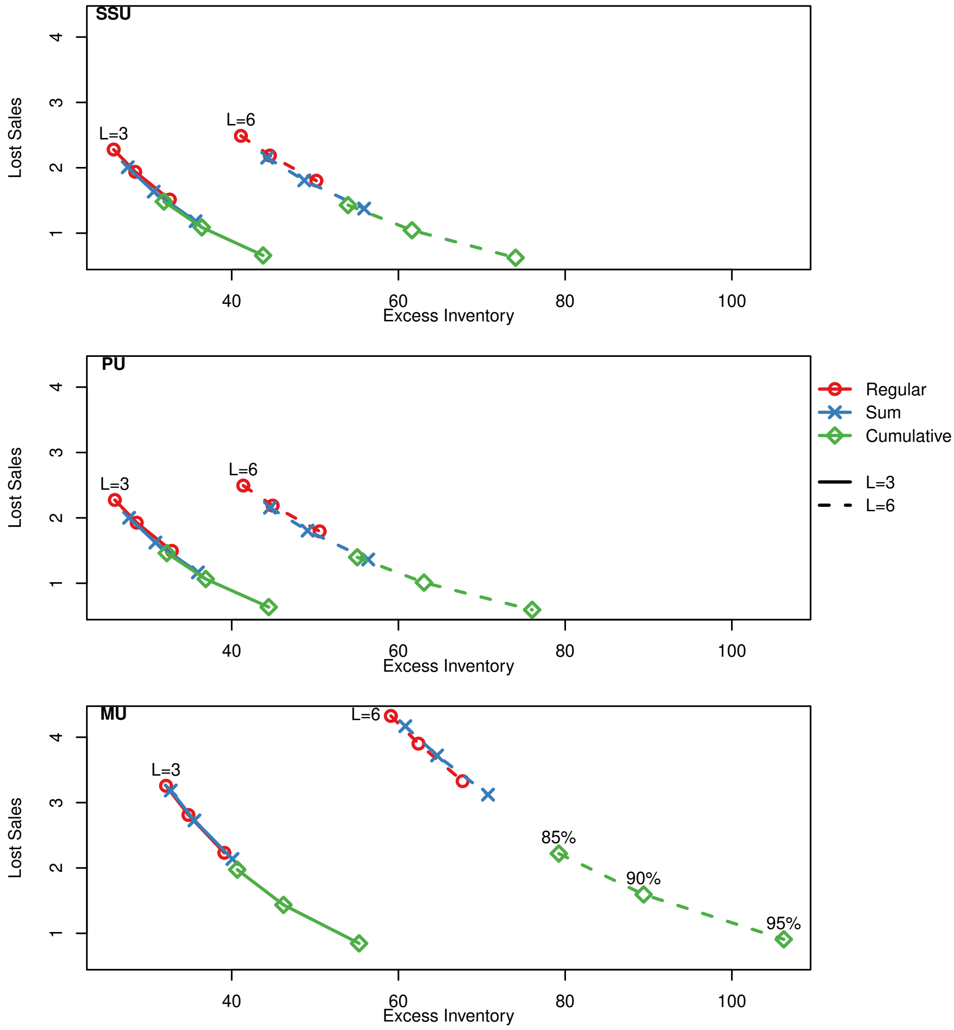

Figure 1: Trade-off curves for stationary demand processes (by lead time and uncertainty type). The values displayed on the vertical axis denote the total lost sales over the test set, while those on the horizontal axis represent the total excess inventory for the same period. The correspondence of the curves to the target service levels (85%, 90% and 95%) is provided for one instance for the curve on the right, corresponding to cumulative approach with lead time 6.

right, which implies that more excess inventory is incurred (which is to be expected as the lead time increases). In addition, we also notice an upward shift from $L + R = 3$ to $L+R = 6$ in lost sales, which leads us to conclude that both approximations display a poor performance. On the other hand, the cumulative approach does not exhibit a substantial increase in lost sales, demonstrating the need to account for the covariances discussed in the theoretical part.

Overall, we find that across all cases, i.e. process type, uncertainty type and lead time, on average the cumulative approach performs best, followed by the sum of variance and the regular ranking the worst. The sum of variances captures better the long-term forecast error variances, but ignores the additional covariance terms that are reflected in the calculation procedure of the cumulative approach. The latter exhibits the minimal deviation from the target service level, which are consistently small, across all cases.

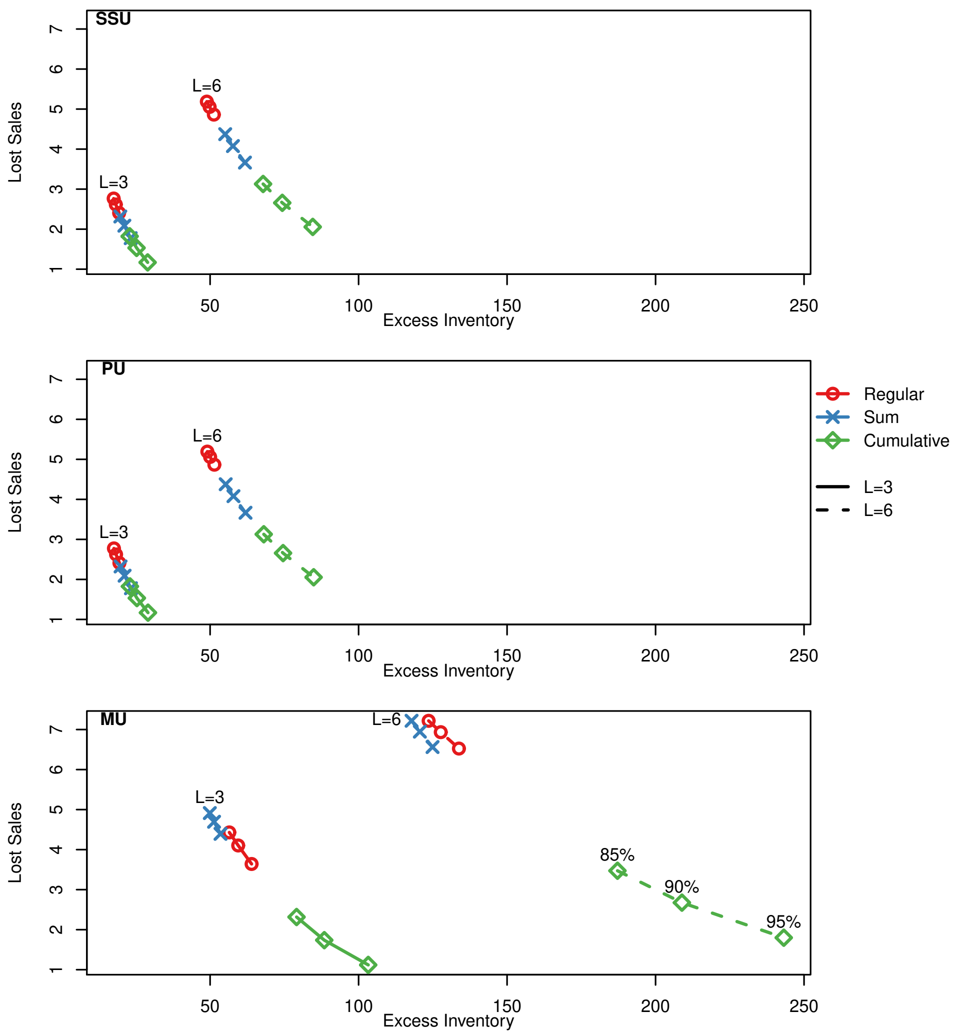

Figure 2: Trade-off curves for non-stationary demand processes (by lead time and uncertainty type). The correspondence of the curves to the target service levels (85%, 90% and 95%) is provided for one instance.

This is achieved without shifting to a dominated trade-off curve, that is it does not need to accumulate unreasonably excess stock and simply results in an appropriate balance between lost sales and excess stock. The other approaches do not achieve this due to the omitted covariances. Its superior performance becomes even more attractive given its implementation simplicity.

6. Real Data Study

We validate the simulation insights by testing the competing variance approximation methods on a real dataset. We use sales data of dairy and cheese products from an American retailer, containing 111 daily series ranging from 521 to 2860 observations.[1]

[1]The dataset used is that of *Dominick's Finer Food*, made publicly available by the University of Chicago at : http://research.chicagogsb.edu/marketing/databases/dominicks/

Compared to the simulated DGPs, many of these time series exhibit promotional effects and seasonality, which would be particularly prone to Model Uncertainty. Similar to before, we use an (R, S) inventory policy, with the same service levels and lead times, such as L > R, and a threefold partitioning of the data. The "burn-in" period and test sets consist of 75 observations each, and the remainder is used to estimate the forecasting model parameters. This can be any exponential smoothing model, as selected by minimising the AICc criterion (for details see Hyndman and Khandakar, 2008). Naturally, even with model selection, this setting falls under the model uncertainty case, as the true underlying data process is unknown. We acknowledge that in practice the retailer might employ a different inventory policy, given software limitations, the nature of the product and the sampling frequency of the data collected; however the use of this dataset is dictated by our experimental requirements rather than mirroring the retailing environment.

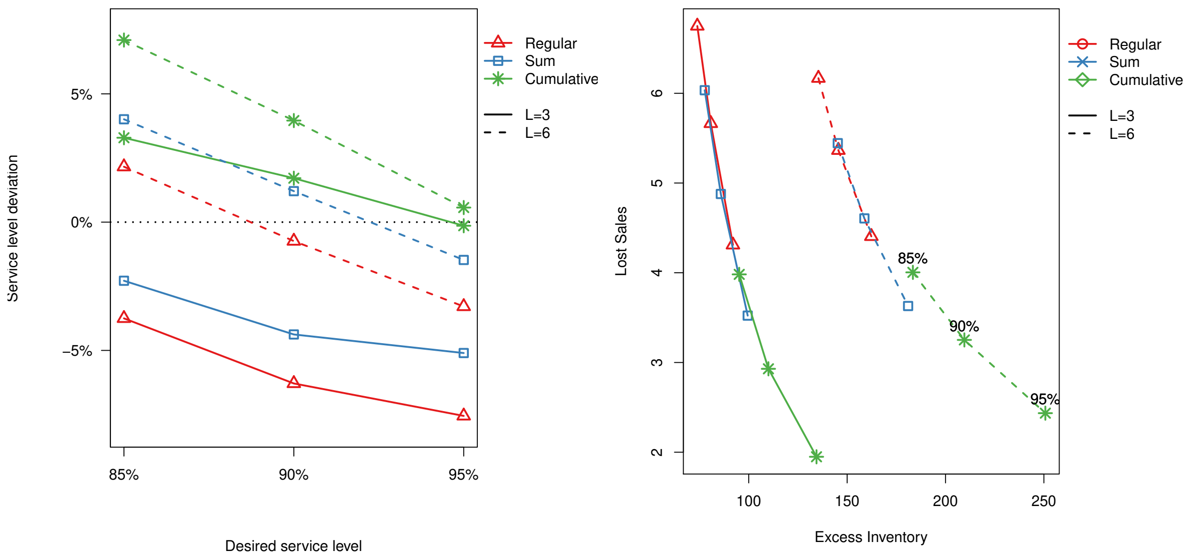

Figure 3: α-Service levels deviations (left) and trade-off curves (right) for the case study data

The results are presented in Figure 3 that plots the service level deviations and trade-off curves. The service level deviations range lies between $[-7.55\%, 7.1\%]$. A tendency to over-cover can be observed, especially for the lowest service level of 85% or high lead times, which can be attributed to the presence of promotions that result in higher forecast error variance estimates, thus further exacerbating the model uncertainty component. Overall, the performance of the cumulative errors is better than the other methods. This is particularly evident for the 95% service level, where its deviations lie very closely to the optimal deviation line, as opposed to the other two methods that return higher deviations.

The trade-off curves in Figure 3 reveal a similar story to those of the simulated data, as the cumulative errors return lower levels of lost sales at the cost of more stock on hand. Again, similarly to the simulation findings, no approximation dominates both aspects of minimising lost sales and excess inventory. But the important result from the graph is the progression of the deviations as the service level increases. Indeed, at higher service levels (which are the norm in practice), the cumulative approach tends to get close to the dashed horizontal line corresponding to zero deviations, which is the ideal scenario, while the other two methods, as they have negative deviations, further diverge from the zero line. The regular approximation, on the other hand, displays the highest level of stock-outs. No set of curves is clearly dominated by others, and when this is paired with the service level deviations, we conclude that the cumulative variance approximation achieves a preferable trade-off between lost sales and stock on hand.

7. Conclusion

Safety stocks are a key element for many business operations, as they allow decision makers to alleviate the effect of demand uncertainty on order levels. A critical component of safety stocks lies in the estimation of the variance of forecast errors, which is used to quantify the uncertainty surrounding it. When computing this variance, the correlation between forecast errors at different lead times is often ignored, which results in lower estimates, which in turn generate sub-par ordering decisions. The existence of these correlations stems from the underlying demand model structure and was demonstrated for simple demand processes even under full knowledge of the generating scheme. This paper has examined this issue in a twofold fashion by: (i) pointing out the shortcomings of the standard heuristic employed by many, and (ii) arguing for the use of cumulative errors, an easy and straightforward empirical approximation, which takes advantage of the distribution of the forecast errors over lead time. The theoretical expressions demonstrated that for the demand processes investigated, the $L\widehat{\sigma}^2_{t+1|t}$ method is not suited to approximate the variance of demand, as it will understate it. This was investigated by deriving the exact lead time forecast error variances of simple ARIMA models, and comparing these expressions with the standard approximation. The Monte Carlo simulation indicated the superiority of the proposed heuristic over the traditional one for an Order-Up-To inventory policy with deterministic lead times. These results held for three types of demand uncertainty. Furthermore, cumulative errors seemed to handle model uncertainty, the typical case in practice, better than its counterparts. These findings were supported by similar ones for a case study using a real dataset from a retailer.

We argue that the recommendations of the paper are directly relevant to practice. Any benefits observed did not come from changes in the forecasting models, or the inventory policy, but rather from how the existing information on forecast errors was used. This is a relatively easy change for existing systems, requiring limited effort with observable results in the outcomes of inventory management. Irrespective of the quality of the forecasts, using cumulative errors will provide benefits, and does not assume unrealistically knowledge of the underlying demand process. Naturally, if this is paired with improving forecasting, we anticipate additional benefits.

There are cases we anticipate this will not perform well, or potentially have no impact. Constructing a distribution of cumulative errors requires more sample than the cruder alternative approximations discussed here. The amount of additional data depends on the relevant lead times, but it is safe to say that when there is a very limited sample the estimation error of the recommended approach can exceed any benefits. This can be relevant for new products that do not have sufficient sample yet. In Section 3 we also showed cases where the impact of the covariances was minimal. Although in these cases one could argue that the effort of switching to cumulative errors may outweigh any benefits, we question how realistic this is. Real time series typically seem to follow sufficiently complex data generating processes, which are further contaminated by disruptions and special events. Nonetheless, the clear disadvantage of all empirical approaches is that they may introduce difficulties to operationalise that theoretical ones can avoid. We call for further research on this, in particular as there is an increasing use of machine learning to generate forecasts that makes a theoretical treatment very difficult.

This paper has shown that using cumulative errors, while being an empirical approximation, is appropriate for finding the variance of forecast errors over lead time. While not being a new method, this work provides the motivation to prefer it over more common approaches and provides an inventory based evaluation of its performance. Nonetheless, in our evaluation we considered the simple Order-Up-To policy. Further research should investigate the effects of the discussed covariances with more involved inventory policies.

A. Derivation of the Covariance Expressions

Recall that $\text{Cov}(aU + bV, cY + dZ) = ac\text{Cov}(U, Y) + bc\text{Cov}(U, Y) + ad\text{Cov}(V, Y) + bd\text{Cov}(V,Z)$, with a, b, c, d constants, and U, V, Y, Z variables. Given

$$\begin{cases} \text{Cov}(\varepsilon_{t+i}, \varepsilon_{t+i}) = \text{Var}(\varepsilon_{t+i}) = \sigma^2 \\ \text{Cov}(\varepsilon_{t+i}, \varepsilon_{t+j}) = 0, \quad i \neq j, \end{cases}$$

and since the error $e_{t+L|t}$ is a linear function of the model innovations up to lead time L, the covariances can be computed by expanding the former in order for the innovations terms to appear in the estimation.

A.1. ARIMA(0,1,0)

For any integers J and K, such that $J < K$, we have

$$e_{t+K|t} = \sum_{i=1}^{K} \varepsilon_{t+i} = \sum_{i=1}^{J} \varepsilon_{t+i} + \sum_{i=J+1}^{K} \varepsilon_{t+i} = e_{t+J|t} + \sum_{i=J+1}^{K} \varepsilon_{t+i}. \tag{A.1}$$

Because of the model structure, $e_{t+K|t}$ can be written as a sum of $e_{t+J|t}$ and other succeeding innovations. For $J < K$, we have

$$
\begin{aligned}
\text{Cov}\left(e_{t+J|t}, e_{t+K|t}\right) &= \text{Cov}\left(e_{t+J|t}, e_{t+J|t} + \sum_{i=J+1}^{K} \varepsilon_{t+i}\right) \\
&= \text{Cov}\left(e_{t+J|t}, e_{t+J|t}\right) + \sum_{i=J+1}^{L} \text{Cov}\left(e_{t+J|t}, \varepsilon_{t+i}\right) \\
&= Var\left(e_{t+J|t}\right) + 0 \\
&= \text{Var}\left(\sum_{i=1}^{J} \varepsilon_{t+i}\right) \\
&= J\sigma^2
\end{aligned} \tag{A.2}
$$

A.2. ARIMA(0,0,1)

For two integers $J < K$, we have:

$$
\begin{aligned}
\text{Cov}(e_{t+J|t}, e_{t+K|t}) &= \text{Cov}(\varepsilon_{t+J} + \theta\varepsilon_{t+J-1}, \varepsilon_{t+K} + \theta\varepsilon_{t+K-1}) \\
&= \text{Cov}(\varepsilon_{t+J}, \varepsilon_{t+K}) + \theta\text{Cov}(\varepsilon_{t+J-1}, \varepsilon_{t+K}) \\
&+ \theta\text{Cov}(\varepsilon_{t+K-1}, \varepsilon_{t+J}) + \theta^2\text{Cov}(\varepsilon_{t+K-1}, \varepsilon_{t+J-1}).
\end{aligned} \tag{A.3}
$$

If $K \neq J + 1$ then the covariance terms all reduce to 0. If however $K = J + 1$, then

$$
\begin{aligned}
\text{Cov}(e_{t+J+1|t}, e_{t+J|t}) &= \text{Cov}(\varepsilon_{t+J+1}, \varepsilon_{t+J}) + \theta\text{Cov}(\varepsilon_{t+J+1}, \varepsilon_{t+J-1}) \\
&+ \theta\text{Cov}(\varepsilon_{t+J+1-1}, \varepsilon_{t+J}) + \theta^2\text{Cov}(\varepsilon_{t+J-1}, \varepsilon_{t+J-1}) \\
&= \theta\text{Var}(\varepsilon_{t+J+1-1}) \\
&= \theta\text{Var}(\varepsilon_{t+J}) \\
&= \theta\sigma^2
\end{aligned} \tag{A.4}
$$

A.3. ARIMA(0,1,1)

For two integers $J < K$, $e_{t+K|t} = \theta\varepsilon_t + \sum_{i=1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K}$ $e_{t+J|t} = \theta\varepsilon_t + \sum_{i=1}^{J-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+J}$, and $e_{t+K|t} = \theta\varepsilon_t + \sum_{i=1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K}$. This implies that

$$
\begin{aligned}
e_{t+K|t} &= \theta\varepsilon_t + \sum_{i=1}^{J}(1+\theta)\varepsilon_{t+i} + \sum_{i=J+1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K} \\
&= \theta\varepsilon_t + \sum_{i=1}^{J-1}(1+\theta)\varepsilon_{t+i} + (1+\theta)\varepsilon_{t+J} + \sum_{i=J+1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K} \\
&= e_{t+J|t} + \theta\varepsilon_{t+J} + \sum_{i=J+1}^{K-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+K}
\end{aligned} \tag{A.5}
$$

The covariance between the errors at these lead times is

$$\text{Cov}(e_{t+J|t}, e_{t+K|t}) = \text{Cov}\left(e_{t+J|t}, e_{t+J|t} + \theta\varepsilon_{t+J} + \sum_{i=J+1}^{K-1} \varepsilon_{t+i} + \varepsilon_{t+K}\right). \tag{A.6}$$

Since $e_{t+J|t}$ is a combination of ε_{t+i} terms up until J, and $J < K$, then $\text{Cov}(e_{t+J|t}, \sum_{i=J+1}^{K-1} \varepsilon_{t+i}) = 0$. The same applies for $\text{Cov}(e_{t+J|t}, \epsilon_{t+K}) = 0$. Therefore, the covariance becomes

$$\begin{aligned}\text{Cov}(e_{t+J|t}, e_{t+K|t}) &= \text{Cov}(e_{t+J|t}, e_{t+J|t} + \theta\varepsilon_{t+J}) \\ &= \text{Cov}(e_{t+J|t}, e_{t+J|t}) + \theta\text{Cov}(e_{t+J|t}, \varepsilon_{t+J}).\end{aligned} \tag{A.7}$$

The second term in the equation amounts to

$$\begin{aligned}\text{Cov}(e_{t+J|t}, \varepsilon_{t+J}) &= \text{Cov}\left(\theta\varepsilon_t + \sum_{i=1}^{J-1}(1+\theta)\varepsilon_{t+i} + \varepsilon_{t+J}, \varepsilon_{t+J}\right) \\ &= \text{Cov}(\varepsilon_{t+J}, \varepsilon_{t+J}) \\ &= \sigma^2.\end{aligned} \tag{A.8}$$

And

$$\begin{aligned}\text{Cov}(e_{t+J|t}, e_{t+J|t}) &= \text{Var}(e_{t+J|t}) \\ &= \sigma^2[(J-1)(1+\theta)^2 + 1],\end{aligned} \tag{A.9}$$

Implying

$$\begin{aligned}\text{Cov}(e_{t+J|t}, e_{t+K|t}) &= \text{Var}(e_{t+J|t}) + \theta\sigma^2 \\ &= \sigma^2[1 + (J-1)(1+\theta)^2 + \theta] \\ &= \sigma^2\left[(1+\theta) + (J-1)(1+\theta)^2\right] \\ &= \sigma^2(1+\theta)\left[1 + (J-1)(1+\theta)\right]\end{aligned} \tag{A.10}$$

This is the covariance between the errors at different lead times.

A.4. ARIMA(1,0,0)

For $J < K$, the error at lead time J is $e_{t+J|t} = \sum_{i=1}^{J} \phi^{(J-i)}\varepsilon_{t+i}$, while that at lead time K is

$$\begin{aligned}e_{t+K|t} &= \sum_{i=1}^{K} \phi^{(K-i)}\varepsilon_{t+i} = \sum_{i=1}^{K} \phi^{(K-J+J-i)}\varepsilon_{t+i} \\ &= \phi^{(K-J)}e_{t+J|t} + \sum_{i=J+1}^{K} \phi^{(K-i)}\varepsilon_{t+i}.\end{aligned} \tag{A.11}$$

The covariance between the errors at these lead times is

$$
\begin{aligned}
\mathrm{Cov}(e_{t+K|t}, e_{t+J|t}) &= \mathrm{Cov}\left(e_{t+J|t}, \phi^{(K-J)}e_{t+J|t} + \sum_{i=J+1}^{K} \phi^{(J-i)}\varepsilon_{t+i}\right) \\
&= \mathrm{Cov}\left(\phi^{(K-J)}e_{t+J|t}, e_{t+J|t}\right) \\
&+ \mathrm{Cov}\left(\phi^{(K-J)}e_{t+J|t}, \sum_{i=J+1}^{K} \phi^{(J-i)}\varepsilon_{t+i}\right) \\
&= \phi^{(K-J)}\mathrm{Var}(e_{t+J|t}) \\
&= \sigma^2\phi^{(K-J)}\left[\frac{1-\phi^{2J}}{1-\phi^2}\right]
\end{aligned} \tag{A.12}
$$

References

Aviv, Y., 2003. A time-series framework for supply-chain inventory management. Operations Research 51 (2), 210–227.

Axsäter, S., 2015. Inventory Control, 3rd Edition. International Series in Operations Research & Management Science. Springer.

Babai, M. Z., Ali, M. M., Boylan, J. E., Syntetos, A. A., 2013. Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: Theory and empirical analysis. International Journal of Production Economics 143 (2), 463–471.

Babai, Z., Dai, Y., Li, Q., Syntetos, A., Wang, X., 2021. Forecasting of lead-time demand variance: implications for safety stock calculations. European Journal of Operational Research.

Badinelli, R. D., 1990. The inventory costs of common mis-specification of demand-forecasting models. International Journal of Production Research 28 (12), 2321–2340.

Barrow, D. K., Kourentzes, N., 2016. Distributions of forecasting error of forecast combinations: Implications for inventory management. International Journal of Production Economics 177, 24–33.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., 2015. Time Series Analysis, Forecasting and Control, 5th Edition. Wiley-Blackwell.

Boylan, J. E., Babai, M. Z., 2016. On the performance of overlapping and non-overlapping temporal demand aggregation approaches. International Journal of Production Economics 181 Part A, 136–144.

Chatfield, C., 1993. Calculating interval forecasts. Journal of Business & Economic Statistics 11 (2), 121–135.

Chatfield, C., 1995. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society. Series A (Statistics in Society) 158 (3), 419–466.

Chatfield, C., Koehler, A. B., 1991. On confusing lead time demand with h-period-ahead forecasts. International Journal of Forecasting 7 (2), 239–240.

Chen, F., Drezner, Z., Ryan, J. K., Simchi-Levi, D., 2000. Quantifying the Bullwhip Effect in a simple supply chain: The impact of forecasting, lead times, and information. Management Science 46 (3), 436–443.

Dong, L., Lee, H. L., 2003. Optimal policies and approximations for a serial multiechelon inventory system with time-correlated demand. Operations Research 51 (6), 969–980.

Eppen, G. D., Martin, R., 1988. Determining safety stock in the presence of stochastic lead time and demand. Management Science 34 (11), 1380–1390.

Fricker, R. D., Goodhart, C. A., 2000. Applying a bootstrap approach for setting reorder points in military supply systems. Naval Research Logistics 47 (6), 459–478.

Gardner, E. S., 1988. A simple method of computing prediction intervals for time series forecasting. Management Science 34 (4), 541–546.

Gardner, E. S., 1990. Evaluating forecast performance in an inventory control system. Management Science 36 (4), 490–499.

Gonçalves, J. N., Carvalho, M. S., Cortez, P., 2020. Operations research models and methods for safety stock determination: A review. Operations Research Perspectives, 100164.

Graves, S. C., 1999. A single-item inventory model for a nonstationary demand process. Manufacturing & Service Operations Management 1 (1), 50–61.

Hyndman, R., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 26 (3), 1–22.

Hyndman, R., Koehler, A., Ord, K., Snyder, R., 2008. Forecasting with Exponential Smoothing: The State Space Approach. Springer.

Janssen, E., Strijbosch, L. W., Brekelmans, R., 2009. Assessing the effects of using demand parameters estimates in inventory control and improving the performance using a correction function. International Journal of Production Economics 118 (1), 34–42.

Johnston, F. R., Harrison, P. J., 1986. The variance of lead time demand. Journal of the Operational Research Society 37 (3), 303–308.

Kahn, J., 1987. Inventories and the volatility of production. The American Economic Review 77 (4), 667–679.

Kim, H., Ryan, J. K., 2003. The cost impact of using simple forecasting techniques in a supply chain. Naval Research Logistics 50, 388–411.

Koehler, A. B., 1990. An inappropriate prediction interval. International Journal of Forecasting 6 (4), 557–558.

Kourentzes, N., 2013. Intermittent demand forecasts with Neural Networks. International Journal of Production Economics 143 (1), 198–206.

Lee, H. L., Padmanabhan, V., Whang, S., 1997. Information distortion in a supply chain: the Bullwhip Effect. Management Science 43 (4), 546–558.

Lee, Y. S., 2014. A semi-parametric approach for estimating critical fractiles under autocorrelated demand. European Journal of Operational Research 234, 163–173.

Lee, Y. S., Scholtes, S., 2014. Empirical prediction intervals revisited. International Journal of Forecasting 30 (2), 217–234.

Makridakis, S., Winkler, R. L., 1989. Sampling distributions of post-sample forecasting errors. Journal of the Royal Statistical Society. Series C (Applied Statistics), 38 (2), 331–342.

Manary, M. P., Willems, S. P., 2008. Setting safety-stocks targets at Intel in the presence of forecast bias. Interfaces 38 (2), 112–122.

Manary, M. P., Willems, S. P., Shihata, A. F., 2009. Correcting heterogeneous and biased forecast error at Intel for supply chain optimization. Interfaces 39 (5), 415–427.

Prak, D., Teunter, R., 2019. A general method for addressing forecasting uncertainty in inventory models. International Journal of Forecasting 35 (1), 224–238.

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

Prak, D., Teunter, R., Syntetos, A., 2017. On the calculation of safety stocks when demand is forecasted. European Journal of Operational Research 256, 454–461.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Ray, W. D., 1982. ARIMA forecasting models in inventory control. Journal of the Operational Research Society 33 (6), 567–574.

Ritchken, P. H., Sankar, R., 1984. The effect of estimation risk in establishing safety stock levels in an inventory model. Journal of the Operational Research Society 35 (12), 1091–1099.

Saldanha, J., Price, B., Thomas, D. J., 2020. A non-parametric approach for setting safety stock levels. Available at SSRN 3624998.

Silver, E. A., Pyke, D. F., Thomas, D. J., 2016. Inventory and Production Management in Supply Chains, 4th Edition. CRC Press.

Snyder, R. D., Koehler, A. B., Hyndman, R. J., Ord, J. K., 2004. Exponential Smoothing models: Means and variance for lead-time demand. European Journal of Operational Research 158, 444–455.

Snyder, R. D., Koehler, A. B., Ord, J. K., 2002. Forecasting for inventory control with exponential smoothing. International Journal of Forecasting 18, 5–18.

Stößlein, M., Kanet, J. J., Gorman, M., Minner, S., 2014. Time-phased safety stocks planning and its financial impacts: Empirical evidence based on European econometric data. International Journal of Production Economics 149, 47–55.

Strijbosch, L. W., Syntetos, A. A., Boylan, J. E., Janssen, E., 2011. On the interaction between forecasting and stock control: The case of non-stationary demand. International Journal of Production Economics 133, 470–480.

Syntetos, A. A., Boylan, J. E., 2006. On the stock control performance of intermittent demand estimators. International Journal of Production Economics 103 (1), 36–47.

Tashman, L. J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. International journal of forecasting 16 (4), 437–450.

Taylor, J. W., Bunn, D. W., 1999. A quantile regression approach to generating prediction intervals. Management Science 45 (2), 225–237.

Thombs, L. A., Schucany, W. R., 1990. Bootstrap prediction intervals for autoregression. Journal of the American Statistical Association 85 (410).

Trapero, J. R., Cardós, M., Kourentzes, N., 2019a. Empirical safety stock estimation based on Kernel and GARCH models. Omega 84, 199–211.

Trapero, J. R., Cardós, M., Kourentzes, N., 2019b. Quantile forecast optimal combination to enhance safety stock estimation. International Journal of Forecasting 35 (1), 239–250.

Trapero, J. R., Cardï¿½s, M., Kourentzes, N., 2018. Empirical safety stock estimation based on kernel and garch models. Omega.

Urban, T. L., 2005. A periodic-review model with serially-correlated, inventory-level-dependent demand. International Journal of Production Economics 95 (3), 287–295.

Wagner, H. M., 2002. And then there were none. Operations Research 50 (1), 217–226.

Wang, M.-C., Rao, S., 1992. Estimating reorder points and other management science applications by bootstrap procedure. European Journal of Operational Research 56, 332–342.

Zhang, X., 2007. Inventory control under temporal demand heteroscedasticity. European Journal of Operational Research 182, 127–144.