Empirical safety stock estimation based on kernel and GARCH models

Juan R. Trapero[a,*], Manuel Cardós[b], Nikolaos Kourentzes[c]

[a]*Universidad de Castilla-La Mancha*
Department of Business Administration, Ciudad Real 13071, Spain
[b]*Universidad Politécnica de Valencia*
Department of Business Administration, Valencia 46022, Spain
[c]*Lancaster University Management School*
Department of Management Science, Lancaster, LA1 4YX, UK

Abstract

Supply chain risk management has drawn the attention of practitioners and academics alike. One source of risk is demand uncertainty. Demand forecasting and safety stock levels are employed to address this risk. Most previous work has focused on point demand forecasting, given that the forecast errors satisfy the typical normal i.i.d. assumption. However, the real demand for products is difficult to forecast accurately, which means that—at minimum—the i.i.d. assumption should be questioned. This work analyzes the effects of possible deviations from the i.i.d. assumption and proposes empirical methods based on kernel density estimation (non-parametric) and GARCH(1,1) models (parametric), among others, for computing the safety stock levels. The results suggest that for lower lead times, normality deviation is more important, and kernel density estimation is most suitable. By contrast, for longer lead times, GARCH models are more appropriate because the autocorrelation of the variance of the forecast errors is the most important deviation. In fact, such autocorrelation can be present, even when no autocorrelation is present in the original demand, as a consequence of the overlapping process used to compute the lead time forecasts and the uncertainties arising in the estimation of the forecasting model parameters. Improvements are shown

*Corresponding author.
Email addresses: `juanramon.trapero@uclm.es` (Juan R. Trapero), `mcardos@doe.upv.es` (Manuel Cardós), `nikolaos@kourentzes.com` (Nikolaos Kourentzes)

in terms of cycle service level, inventory investment and backorder volume. Simulations and real demand data from a manufacturer are used to illustrate our methodology.

Keywords: Forecasting, safety stock, risk, supply chain, prediction intervals, volatility, kernel density estimation, GARCH.

1. Introduction

Supply Chain Risk Management (SCRM) is becoming an interesting area for researchers and practitioners, and the number of related publications is growing [1]. A critical review of SCRM is provided by [2]. An important source of risk is the uncertainty of future demand. When demand is unusually large, a stockout may occur, with associated negative consequences. When demand is below expectations, the company is saddled with higher holding costs due to excess inventory. To mitigate such risks, companies use safety stocks, which are additional units beyond the stock required to meet a lead time forecasted demand. Different approaches can be utilized to calculate the safety stock [3]. Although the most appropriate method depends on an organization's circumstances, calculating the safety stock based on customer service, which is determined from demand uncertainty, is widely used, because it does not require knowledge of the stockout cost, which can be very difficult to estimate.

Demand uncertainty has been estimated by using both demand variability [4] and demand forecast error variability ([3], [5]). In this work, we follow the latter approach, given that future demand is unknown and must be forecasted. Therefore, safety stocks are intended to prevent issues due to such demand forecast errors.

Demand forecast errors are typically assumed to be independent and identically distributed (i.i.d.), following a Gaussian distribution with a zero mean and a constant variance. However, the forecast error often does not satisfy these assumptions, which can cause the achieved service levels to deviate from the target levels, with increased inventory costs [6, 7, 8, 9]. Even when the distribution assumptions are met, the uncertainties that result from estimating the forecasting models are usually not considered [10, 11].

Two approaches can be used to address the problem of forecast errors that do not fulfill i.i.d. assumptions. First, a theoretical approach, in which the demand forecasting model should be improved by modeling, for example, mean demand autocorrelation [12, 13]; including exogenous variables such as promotional variables [14, 9, 15] or weather variables [9] via regression-type models; as well as considering temporal heteroscedastic processes [16, 17, 18, 9, 19]. Under the assumption that the improved forecasting model captures the underlying data process, a theoretical formula for the cumulative variance of the forecast error can be determined [20]. Then, the safety stock can be computed using that theoretical formula and assuming a statistical distribution.

Nonetheless, it would be unrealistic to assume that we could find the true demand model for each SKU, and even if such a model did exist, it would not always be possible for companies to implement it in practice. For instance, many companies judgmentally adjust statistical forecasts, which may induce forecast error biases [21, 22]. Furthermore, as pointed out by Fildes [23], companies often do not implement the latest research in forecasting, relying instead on software vendors who have themselves been shown to be very conservative. Moreover, the choice of demand forecasting model may not be under the control of the operations/inventory planning department [6].

For the aforementioned reasons, an alternative to the theoretical approach is a data-driven empirical counterpart. In a univariate framework, the empirical approach involves collecting the observed lead time forecast errors of whatever forecasting model a company has available and, subsequently, obtaining the quantiles of interest to determine the safety stock. Note that in this empirical approach, we need to know neither the point forecasting model nor its parameters. This is potentially advantageous in the case of forecasting support systems, which often do not provide such information to users. For the case in which exogenous variables are available, Beutel and Minner [9], working within a multivariate framework, proposed a data-driven linear programming approach in which the inventory level could be optimized by minimizing a cost objective or subject to a service level constraint.

Depending on which i.i.d. assumptions are violated, different options can be applied to improve the empirical safety stock calculation. For a multivariate case, Beutel and Minner [9] showed that the linear programming approach was robust to unconditional both bias (where demand is assumed to be independent of time) and heteroscedasticity as well as non-normal residuals. In the same reference, it was reported that in the case of heteroscedasticity due to independent variables, a multiple linear regression optimized via the ordinary least squares approach could be improved by relating standard deviations to the levels of explanatory variables (for example, price) in a parametric functional form.

In a univariate framework, several methodologies have been proposed to correct the problem that arises in this case if the lead time forecast errors are biased. The authors of [6] adjusted the standard deviation of the forecast errors and reference [24] proposed a kernel-smoothing technique to cope with such unconditional bias. If the mean demand is autocorrelated and it is not properly modeled, then the forecast error bias (if any) can be autocorrelated as well. Lee [8] developed a semi-parametric approach to estimate the critical fractiles for autocorrelated demand, where the conditional lead time demand forecast bias is a parametric linear function of the recent demand history and the quantile is obtained non-parametrically from the empirical distribution of the regression errors.

Likewise, lead time forecast errors may be heteroscedastic and deviate from normality [24, 8]. Non-normal residuals have also been addressed by kernel smoothing [24] and the empirical lead time error distribution quantile approach [8]. However, the case when the lead time forecast errors exhibit conditional heteroscedasticity remains understudied.

In this work, the goal is to compute the safety stock for a given lead time using empirical methods on the measured lead time forecast errors. Generalised AutoRegressive Conditional Heteroscedastic (GARCH) models [25] and exponential smoothing models [26] will be investigated to overcome the assumption of the independence of lead time forecast error standard deviation and exploit potential correlations.

To the best of the authors' knowledge this is the first time that GARCH models have been applied empirically to forecast lead time error standard deviations and compute safety stocks. Furthermore, we will explore the use of empirical non-parametric approaches such as kernel density estimation [27], which is typically utilized in prediction interval studies [28], although its use in supply chain applications has been less frequent [24, 29]. The proposed empirical approaches will be also compared with the traditional supply chain theoretical approaches, which are based on i.i.d. assumptions, using both simulated data and real data from a manufacturing company. The performance of the different approaches will be measured using inventory investment, service level and backorders metrics. First, a newsvendor-type model will be assumed, and subsequently, the results will be compared with an order-up-to level stock control policy.

The remainder of the paper is organized as follows: Section 2 formulates the safety stock problem. Section 3 reviews the theoretical approach for computing the safety stock. Section 4 describes the empirical approximation proposed in this work. Section 5 explains the performance metrics that will be used to assess the different methods. This section also provides implementation details for the alternative approaches. Simulation experiments are carried out in Section 6. A case study based on shipment data from a real manufacturer is provided in Section 7. Finally, Section 8 presents concluding remarks.

2. Problem formulation

When the demand forecasting error is Gaussian i.i.d. with zero mean and constant variance, the safety stock (SS) for a target Cycle Service Level (CSL), expressed as the target probability of no stockout during the lead time, can be computed as follows:

$$SS = k\sigma_L, \tag{1}$$

where $k = \Phi^{-1}(CSL)$ is the safety factor, $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and σ_L denotes the standard deviation of the forecast error for a certain lead time L, which is assumed to be constant and known.

The main problem in (1) is how to estimate σ_L. To do that, we consider two approaches: a theoretical approach based on a forecasting model, in which an estimate of σ_1 (one-step-ahead standard deviation of the forecast error) is provided and an analytical expression that relates σ_L to the forecasting model parameters, the lead time and σ_1 is employed. The estimation of σ_1 is possible because the forecast updating step is usually smaller than the lead time. Alternatively, an empirical parametric approach can also be employed in which σ_L is estimated directly from the lead time forecast error. Here, the term parametric indicates that a known statistical distribution is assumed; normality is traditionally chosen for this purpose.

If the statistical distribution of the error is unknown, empirical non-parametric methods can be used instead. In that case, the safety stock calculation should be reformulated as follows:

$$SS = Q_L(CSL), \tag{2}$$

where $Q_L(CSL)$ is the forecast error quantile at the probability defined by CSL. We discuss the theoretical approach in detail below and the empirical parametric and non-parametric approaches in the subsequent section.

3. Theoretical approach

3.1. Estimation of σ_1

Traditional textbooks suggest computing σ_1 based on forecast error metrics. For instance, [3] uses Mean Squared Error (MSE) and [5] uses Mean Absolute Deviation (MAD). However, the conversion factor from MAD to σ_1 depends on the assumed statistical distribution and an inappropriate choice can result in inadequate safety stock to meet the required service level [3]. Therefore, hereafter, we use MSE to estimate σ_1, such as $\sigma_1 = \sqrt{MSE}$. Moreover, the MSE estimation can be updated as new observations become available, for example:

$$MSE_{t+1} = \alpha'(y_t - F_t)^2 + (1 - \alpha')MSE_t, \tag{3}$$

where y_t is the actual value at time t, F_t is the forecast value for the same time period, MSE_{t+1} is the MSE forecast for period $t+1$ and α' is a smoothing constant that varies between 0 and 1. In this work, α' and the initial value are optimized by minimizing the in-sample squared error following the suggestion by [30]. Note that (3) is the well-known Single Exponential Smoothing (SES) method.

3.2. Estimation of σ_L

A theoretical formula for σ_L can be obtained given the following: σ_1; the forecast updating procedure; the value of the lead time (L), which is assumed to be constant and known; and the values of the smoothing constants used. According to [3], the exact relationship can be complicated. If we disregard the fact that the errors usually increase for longer forecast horizons and also assume that the forecast errors are independent over time [31], then

$$\sigma_L = \sqrt{L}\sigma_1. \tag{4}$$

It should be noted that expression (4) has been seriously criticized in [20] because there is no theoretical justification for its use and it can produce very inadequate results.

The authors of [32, 33, 34, 35] showed that when the demand can be modeled using a local level model (i.e., the model that underlies SES) with the parameter α, the conditional variance for the lead time demand is

$$\sigma_L = \sigma_1\sqrt{L}\sqrt{1 + \alpha(L-1) + \frac{1}{6}\alpha^2(L-1)(2L-1)}. \tag{5}$$

Interestingly, although (5) provides an exact relationship, the literature often relies on the approximation in (4) (for example, [6, 7]).

4. Empirical approach

The theoretical approach represented in (5) assumes that forecast errors are i.i.d., which means that we know the "true" model of the SKU's demand. However, if there are doubts about the validity of the "true" model, then empirical approaches can be a useful alternative [20]. Given the complex nature of markets, clients, promotions, economic situation and so forth, assuming that we know the true demand model for each SKU would be unrealistic; consequently, empirical approaches must be, at least, tested.

Note that although empirical approaches have been shown to yield good results in other applications, such as the calculation of prediction intervals [20], they have rarely been used in supply chain risk management [8].

4.1. Parametric approach

In these empirical parametric approaches, we retain the assumption that lead time forecast errors are normally distributed, but we relax the variance independence assumption by allowing σ_L to vary over time. An empirical estimate of σ_L can be calculated as follows:

$$\sigma_L = \sqrt{\frac{\sum_{t=1}^{n}(e_t(L) - \bar{e}(L))^2}{n}}, \tag{6}$$

where $e_t(L) = y_L - F_L = \sum_{h=1}^{L} y_{t+h} - \sum_{h=1}^{L} F_{t+h}$ is the lead time forecast error and $\bar{e}(L)$ is the average error for the L under consideration.

4.1.1. Single Exponential Smoothing (SES)

Instead of using (6) to update σ_L each time a new observation is available, we can use SES to directly update the cumulative MSE over lead time [36, 26]. Likewise, we can update σ_L on the basis of the lead time forecast error instead of the one-step-ahead forecasting error:

$$MSE_{L,t+1} = \alpha''(y_L - F_L)^2 + (1 - \alpha'')MSE_{L,t}, \tag{7}$$

where $\sqrt{MSE_{L,t+1}}$ forecasts σ_L at time $t+1$. Unlike Eq. (3), the updating step does not match with the lead time forecast error. The expression is updated after every period t. The process followed to calculate y_L and F_L is known as the overlapping temporal demand aggregation process [37], which the authors recommend when sufficient demand history is available (at least, 24 observations). Similar to Eq. (3), both α'' and the initial value for Eq. (7) are optimized by minimizing the squared errors.

4.1.2. ARCH/GARCH models

Although exponential smoothing is the work horse in supply chain forecasting, when we are dealing with volatility forecasting other models have been developed that may be good candidates to enhance risk estimation and, therefore, safety stock determination.

In particular, the AutoRegressive Conditional Heteroscedasticity (ARCH) model introduced by [38] is one of the most important developments in risk modeling, and it may be well-suited for our application. The ARCH model expresses the forecast error as $\epsilon_t = \sigma_t \cdot v_t$, where

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^{p} a_i \epsilon_{t-i+1}^2, \tag{8}$$

where ω is a constant, and a_i contains the respective coefficients of the p lags of v_t that form an i.i.d. process. ARCH models express the conditional variance as a linear function of the p lagged squared error terms. Bollerslev in [25] proposed the generalized autoregressive conditional heteroscedasticity (GARCH) models that represent a more parsimonious and less restrictive version of the ARCH(p) models. GARCH(p,q) models express the conditional variance of the forecast error (or return) (ϵ_t) at time t as a linear function of both q lagged squared error terms and p lagged conditional variance terms. For example, a GARCH(1,1) model is given by:

$$\sigma_{t+1}^2 = \omega + a_1 \epsilon_t^2 + \beta_1 \sigma_t^2. \tag{9}$$

It should be noted that exponential smoothing has the same formulation as the integrated GARCH model (IGARCH) [39], with $\beta_1 = 1 - a_1$ and $\omega = 0$. If we apply the GARCH(1,1) model to lead time forecasting error instead of to the one-step ahead forecasting error, Eq. (9) can be rewritten as follows:

$$\sigma_{L,t+1}^2 = \omega' + a_1' \epsilon_{L,t}^2 + \beta_1' \sigma_{L,t}^2. \tag{10}$$

In this work, we focus our analysis on the GARCH(1,1) model using an overlapping approach to compare it with the SES model shown in Eq. (7). Note that GARCH(1,1) also includes GARCH(0,1).

4.2. Non-parametric approach

Some demand distributions present important asymmetries, particularly when they are subject to promotional periods or special events. In these cases, the typical normality assumption for forecasting errors does not hold. Non-parametric approaches, in which the safety stock is calculated in accordance with (2), are well suited for overcoming this problem. Here, we use two well-known non-parametric methods: (i) kernel density estimation and (ii) empirical percentile.

4.2.1. Kernel density estimation

This technique represents the probability density function $f(x)$ of the lead time forecast errors without requiring assumptions concerning the data distribution. The kernel density formula for a series X at a point x is given by

$$f(x) = \frac{1}{Nh} \sum_{j=1}^{N} K\left(\frac{x - X_j}{h}\right), \tag{11}$$

where N is the sample size, $K(\cdot)$ is the kernel smoothing function that integrates to one and h is the bandwidth [27]. That reference, p. 43, shows that the optimal kernel function, often called the Epanechnikov kernel, is:

$$K_e(t) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}t^2\right) & -\sqrt{5} \leq t \leq \sqrt{5} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Although the choice of h is still a point of debate in statistics literature, typically, the following optimal bandwidth for a Gaussian kernel is chosen [27, 28]:

$$h_{opt} = 0.9AN^{-1/5}, \tag{13}$$

where A is an adaptive estimate of spread given by

$$A = min(\text{standard deviation}, \text{interquantile range}/1.34) \tag{14}$$

The CSL quantile denoted by $Q_L(CSL)$ can be estimated non-parametrically using the empirical distribution fitted by the Kernel approach on the lead time forecast errors.

4.2.2. Empirical percentile

The well-known percentile indicates the value below which a given percentage of observations fall. When calculating empirical percentiles the requested value is typically linearly interpolated from the available values.

5. Evaluation of the alternative approaches

5.1. Point forecast

Before evaluating various approaches for measuring forecast error volatility, we must first define the point forecast function. The point forecast algorithm yields a measure of the central tendency of the forecast density function. This value and the resulting forecast errors are the inputs used for the variability forecasting methods described in Sections 3 and 4. Note that the same point forecast is used in all cases, regardless of the variability forecasting approach considered.

In this work, SES is used to produce point forecasts for two reasons. First, SES is widely used in business applications [40, 41], and it matches common practice, in which a company may not be using the optimal forecasting model due to either lack of expertise or to a limited set of forecasting models available in software. Second, we use SES as a benchmark because an analytical expression is available to compute the lead time forecasting error variance (see Eq. (5)), which requires an estimate of the value of α used in SES-based point forecasting. Such a forecasting method can be formulated as

$$F_{t+h} = \alpha y_t + (1-\alpha)F_t, \tag{15}$$

where $0 \leq \alpha \leq 1$ and h is the forecasting horizon. Given the recursive nature of exponential smoothing, it is necessary to initialize the algorithm. We determine the initial value together with the α value by minimizing the in-sample mean squared error. Note that the lead time demand forecast is $F_L = \sum_{h=1}^{L} F_{t+h} = L \cdot F_{t+1}$.

5.2. Inventory performance metrics

In a supply chain context, according to [42], [43] and [44], trade-off curves are employed to measure the forecasting performance of different techniques. These trade-off curves are the pairs formed by CSL and inventory investment plus backorders and inventory investment. Kourentzes [43] transformed inventory investment and backorder metrics into scale-independent measures that can easily summarize metrics across products by dividing them by the average in-sample demand.

Such inventory metrics are parametrized depending on the stock control policy. Here, we follow a newsvendor-type policy as in [9, 8], where the critical fractile minimizes the expected cost by balancing the costs of understocking and overstocking [31]. In addition, we also studied the critical fractile estimation problem with a fixed replenishment lead time, in which such a critical fractile solution is repeatedly applied for a multi-period inventory model, [8]. The inventory metrics are the achieved CSL, which is estimated as the percentage of iterations where the lead time demand realizations fall below the estimated lead time demand quantile for determined target CSL values (critical fractiles). Backorders, by which we mean the amount of unsatisfied demand, are calculated in two steps. First, we calculate the sum of the demand realizations that exceed the estimated target CSL for each SKU, and second, we calculate the average of that sum across SKUs. Since we are interested in the safety stock, the average inventory investment will be measured as the average of the so-called scaled safety stock, which is the safety stock divided by the in-sample average demand. All three metrics are calculated on the hold-out sample. In section 6.4, a periodic review, order up to level will also be implemented to compare it with the newsvendor policy results. Note that an order-up-to level stock control policy is more appropriate when products are sold over a long time horizon with numerous replenishment opportunities and, thus, there is no need to dispose excess inventory [45].

In this work, the target CSL values are set to 85%, 90%, 95% and 99% for both simulations and real data.

5.3. Implementation of approaches

The analysis was conducted using MATLAB. Here, we briefly list the implementation details of the approaches discussed above. We employed the econometric toolbox from MATLAB to implement the GARCH(1,1) model selecting an interior-point optimization algorithm. We used the kernel density estimators from MATLAB with the default options for the selection of the bandwidth and the Epanechnikov kernel. The empirical percentile was computed using the MATLAB function **prctile**.

6. Simulation results

In this section, we describe the four simulation experiments conducted to explore the performance of the aforementioned empirical parametric and non-parametric approaches when there are deviations from the Gaussian i.i.d. assumptions. We study what happens when: (i) the normality assumption does not hold for different sample sizes, (ii) the homoscedasticity and variance independence assumptions do not hold, and (iii) the lead time varies.

Those simulations will be assessed by means of the newsvendor inventory performance metrics (achieved CSL, scaled safety stock and backorders). To extend these simulation results to another well-known stock control policy, the last simulation (iv) will analyze the relationship between the newsvendor and order-up-to level inventory performance metrics.

In these experiments, we divided the total sample into three parts. We used the first 20% of the data to estimate the parameters (initial value and smoothing parameter) for the point forecast method and the next 50% to estimate the parameters for the volatility forecast methods: SES, kernel density estimation, empirical percentile and GARCH(1,1). Note that empirical methods are estimated using data which have not been employed to produce the point forecast, [46, 47, 48]. The remainder of the data was reserved as a hold-out sample for evaluation purposes.

6.1. Sample size and demand statistical distribution

To explore the influence of demand distributions on the safety stock calculation, we employed three distributions: (i) Normal, (ii) Log-normal and (iii) Gamma. Using this approach, we could observe any issues when the demand was not normal. Note that the log-normal distribution is reasonable when products are subject to promotional periods, during which the observed demand is higher than the baseline demand. Gamma distributions have been used in the literature [9, 29, 49] to assess violations of the normality assumption.

We conducted simulations for the following sample sizes: 50, 100 and 500. Studying the sample size is important because products often have short life cycles; consequently, few historical observations are available. For each sample size and each distribution, we conducted Monte Carlo simulations with 100 repetitions. The simulated population values for the mean (μ) and standard deviation (σ) of the normal distribution were $\mu = 150$ and $\sigma = 25$. The parameters for the log-normal and Gamma distributions were chosen to provide a mean of 150 and a skewness of 9.6, with the latter being used as a non-normality measure. The resulting parameter values for the log-normal distribution were $\mu = 4.7$ and $\sigma = 0.7$, while for the Gamma distribution they were $a = 0.04$ (shape parameter) and $b = 3,449$ (scale parameter).

Figure 1 shows the trade-off curves for a sample size of 50 observations and a lead time of 1. The upper plots show the deviations from the target CSL in percentage vs. the scaled safety stock. For example, with a normal distribution and a target CSL of 85%, denoted by the smallest marker, the CSL deviation for SES is approximately -1% (i.e., undercoverage with an achieved CSL of 84%). The closer the lines are to zero deviation, the better the performance is considered to be. The lower panels plot the relationship between backorders and scaled safety stock. Note that the different target CSLs are organized from the smallest-size marker (85%) to the largest-size marker (99%).

The left column of Figure 1 (both upper and lower panels) shows the trade-off curves for the normally distributed demand. As expected, SES and GARCH perform reasonably well and produce a slight systematic undercoverage, whereas the Kernel method produces a systematic overcoverage that is reduced for the highest target CSL. When the target is 99%, the empirical percentile method achieves a remarkable lower CSL, which also implies a higher volume of backorders.

In the middle column of Figure 1, we find the same trade-off curves for the log-normal distribution. The non-parametric approaches (Kernel and Percentile) achieve the highest and lowest target CSLs (99% and 85%, respectively), whereas the parametric approaches (SES and GARCH) do not. Therefore, when the forecast errors are skewed, the typical normal assumption results in CSL values deviated from their targets. In this sense, the non-parametric approaches seem more robust.

The right column of Figure 1 shows the results for the Gamma-simulated demands. The conclusions here are similar to the log-normal demands, although they are exacerbated. In general terms, the parametric approaches induce overcoverage for lower targets (85%-90%); in fact, the lower the target is, the higher the overcoverage is. Similarly, the higher the target is (95%–99%), the higher the undercoverage is. The empirical non-parametric approaches achieve CSL values that are closer to the targets. This difference between target CSLs is also reflected in the lower plot, where the non-parametric approaches yield high backorder values for lower CSL values and vice versa. By contrast, the parametric approaches yield backorder values

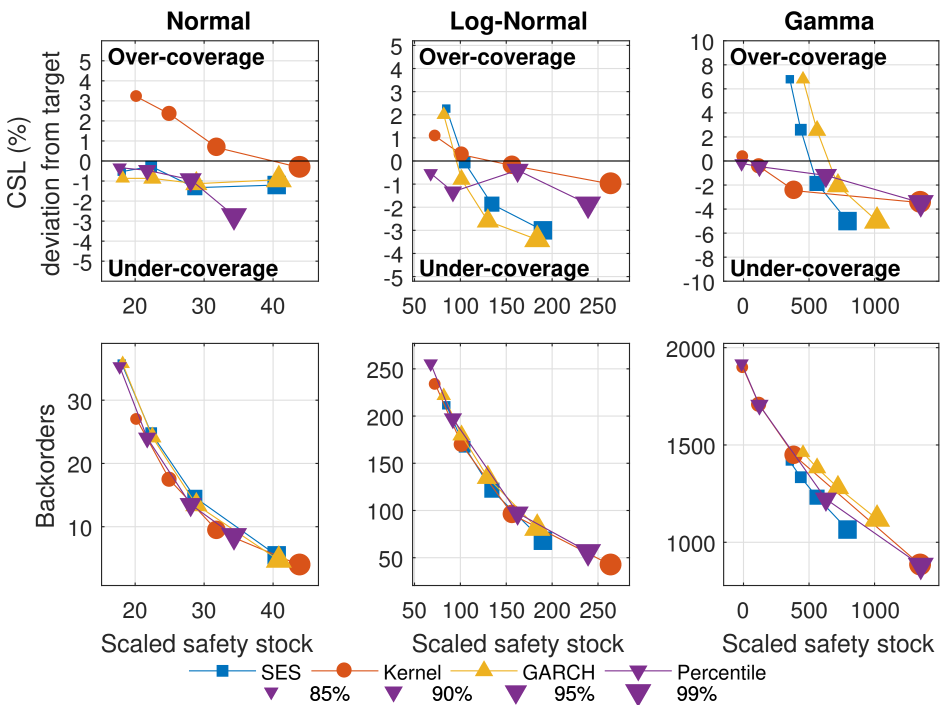

Figure 1: Trade-off curves for a sample size of 50 observations. Left column: Normal distribution; middle column: Log-normal distribution; right column: Gamma distribution.

with lower variations regardless of the CSL value.

Figure 2 depicts the same trade-off curves when the sample size increases to 100 observations. Again, the non-parametric approaches are able to capture the asymmetry of the forecast errors at the highest and lowest CSLs (99% and 85%) for both the log-normal and Gamma simulated demands. When the distribution is normal, all methods show similar performance in achieving values close to the target CSLs, although the Kernel method produces systematically higher service levels than expected.

The same conclusions are drawn when the sample size is increased from 100 to 500 observations.

To summarize, the non-parametric methods such as the Kernel and Percentile methods provide better trade-off curves for the Log-normal and Gamma distributions, since the parametric methods assume that the forecast errors are normal. However, the performance of the non-parametric methods deteriorates when the sample size is small since they are data-driven approaches. When the distribution is normal, the parametric approaches work well, as

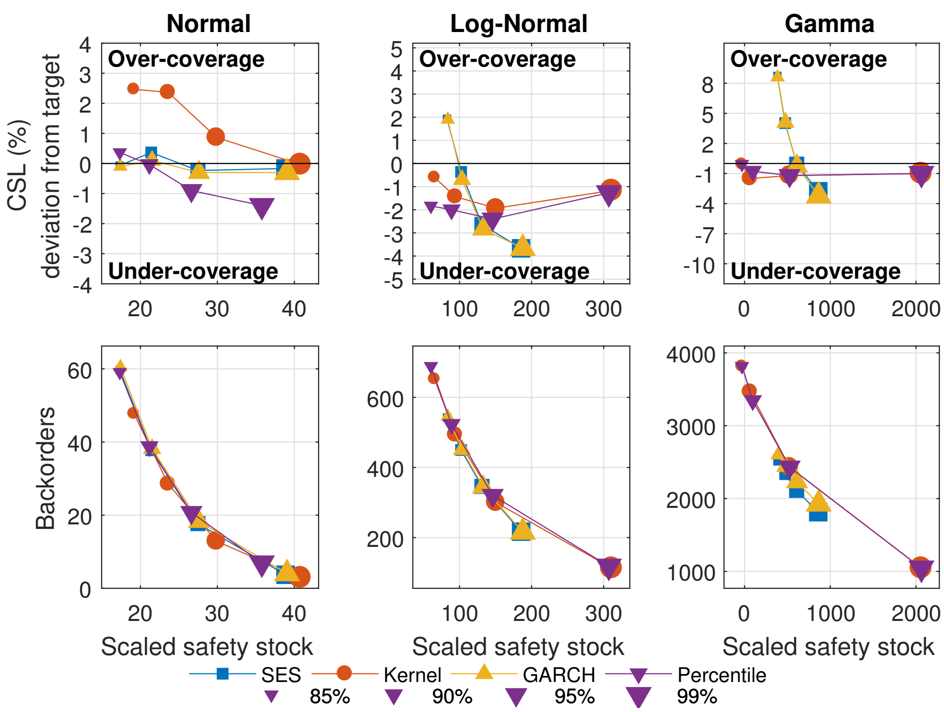

Figure 2: Trade-off curves for a sample size of 100 observations. Left column: Normal distribution; middle column: Log-normal distribution; right column: Gamma distribution.

expected, although when the sample size is small there is an undercoverage because the parameters are not known and have to be estimated [10]. The Kernel method results in overcoverage, and the Percentile method is more sensitive to small sample sizes. The resulting overcoverage of the Kernel method is due to the bandwidth selection, which establishes a compromise between bias and variance [27]. Given that these simulations do not consider correlations of the forecast error variance, the GARCH model does not exhibit any significant advantage over the SES model.

6.2. Demand with time-varying volatility

In this simulation, we conducted two experiments with time-varying volatility. In the first case, the demand followed a normal distribution with a constant mean ($\mu = 150$) and two different standard deviations ($\sigma_1 = 25$ and $\sigma_2 = 50$) with a sample size of 500 observations and a lead time equal to 1. σ_1 was used for the time periods corresponding to 1:50, 101:226 and 351:425. The purpose was to have volatility changes in both the in-samples and hold-out samples. The second case simulated a demand with a constant mean set to 50 with a stochastic term that followed a GARCH(1,1) model, using the parameters ($\omega = 0.01$, $a_1 = 0.4$ and $\beta_1 = 0.5$). Both experiments were repeated 100 times.

Figure 3 depicts the average trade-off curves obtained by each considered technique for both experiments. The right column shows the trade-off curves for the first experiment with volatile abrupt changes. Regarding backorders, the SES and GARCH methods yield slightly lower backorders for similar scaled safety stocks. With respect to CSL deviations, the SES and GARCH methods provide overcoverage for lower targets and undercoverage for the highest target. The Kernel method produces overcoverage for each target, which is reduced for larger target CSLs, whereas the Percentile method achieves a CSL deviation close to zero.

The left column of Figure 3 shows the trade-off curves for the second experiment based on a GARCH simulation. In general terms, the GARCH method produces low CSL deviations with low scaled safety stocks. Although SES achieves backorder values similar to those of GARCH, its CSL performance is much worse than that of either GARCH or the non-parametric approaches. Additionally, the levels of backorders provided by the parametric approaches are lower than those of their non-parametric counterparts, even for lower scaled safety stocks.

These results can be interpreted as follows: (i) When forecast errors have variance autocorrelation, GARCH is a promising alternative that deserves further exploration when computing safety stocks. Thus far, the best option commonly applied was SES [50, 3], which can be seen as a particular case of GARCH models [39]. (ii) These results are also important because they suggest that the level of backorders is a signal of potential autocorrelation in the forecast error variance. In other words, when parametric approaches provide a level of backorders lower than the non-parametric methods, that result can be due to a potential variance autocorrelation. To support this argument Figure 4 shows the actual and point forecasts of a Monte Carlo repetition on the hold-out sample for a CSL target of 95%. The same plot also depicts the safety stock computed by the Kernel and GARCH methods. We can see that GARCH is able to adapt to periods of high/low volatility whereas Kernel is not. Interestingly, although the Kernel and GARCH methods result in similar numbers of stockouts during high-volatility periods and consequently produce similar achieved CSLs, the level of backorders is much higher when the Kernel approach is used. Similarly, during periods of low volatility, the average safety stock yielded by Kernel is greater than that of GARCH, and thus, the inventory investment is also higher with the Kernel approach.

6.3. Influence of lead time

The lead time in the previous simulations was set to 1. In this section, we describe the influence of the lead time on the safety stock computation. For these experiments, we set the lead time to 4 periods, the sample size to 500 observations and each experiment has been repeated 100 times. We used a demand process that followed an ARIMA(0,1,1) model with additional Gaussian innovations of zero mean and a standard deviation of $\sigma = 2$, where $\theta = -0.75$. We used the ARIMA model because SES is optimal for such a model; thus, all differences between the methods considered are independent of the point forecast model. The value of $\theta = -0.75$ corresponds to a theoretical exponential smoothing constant of $\alpha = (1 + \theta) = 0.25$. To ensure positive demand values, a constant of 50 units was added to the signal. Given that the lead time was greater than 1, we can add to the previous empirical approaches the theoretical approaches from Eqs. (4) and (5), hereafter denoted by $\sigma_L(4)$ and $\sigma_L(5)$, respectively. Recall that SES, Kernel, GARCH and Percentile methods are based on empirical lead time forecast errors.

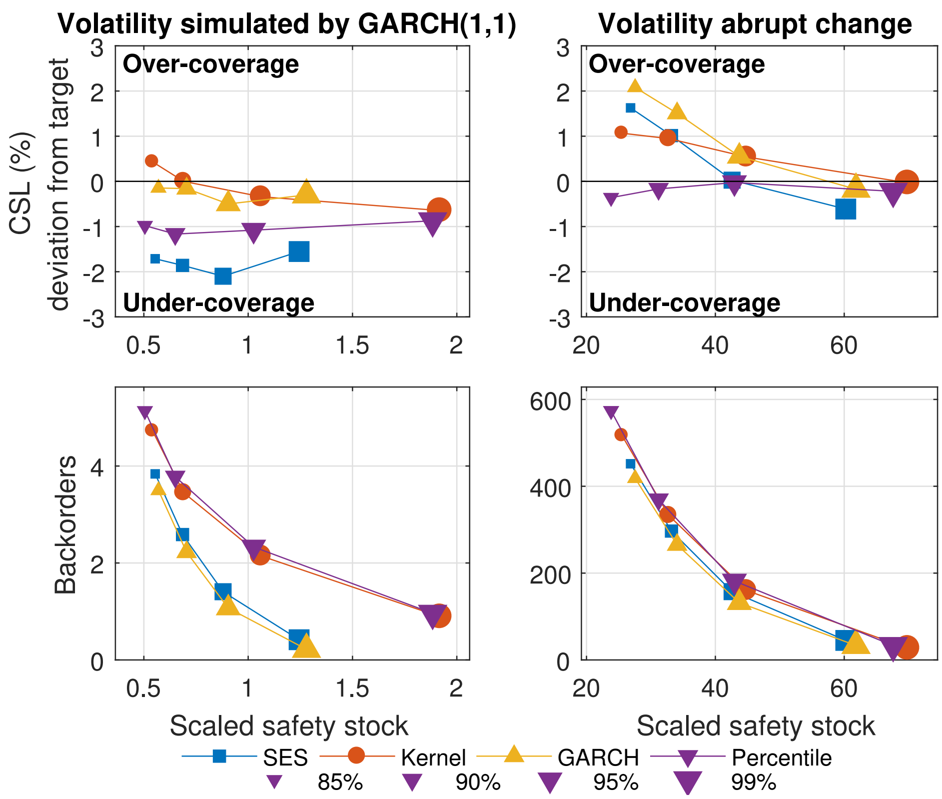

Figure 3: Trade-off curves for the experiments with demand subject to time-varying volatility

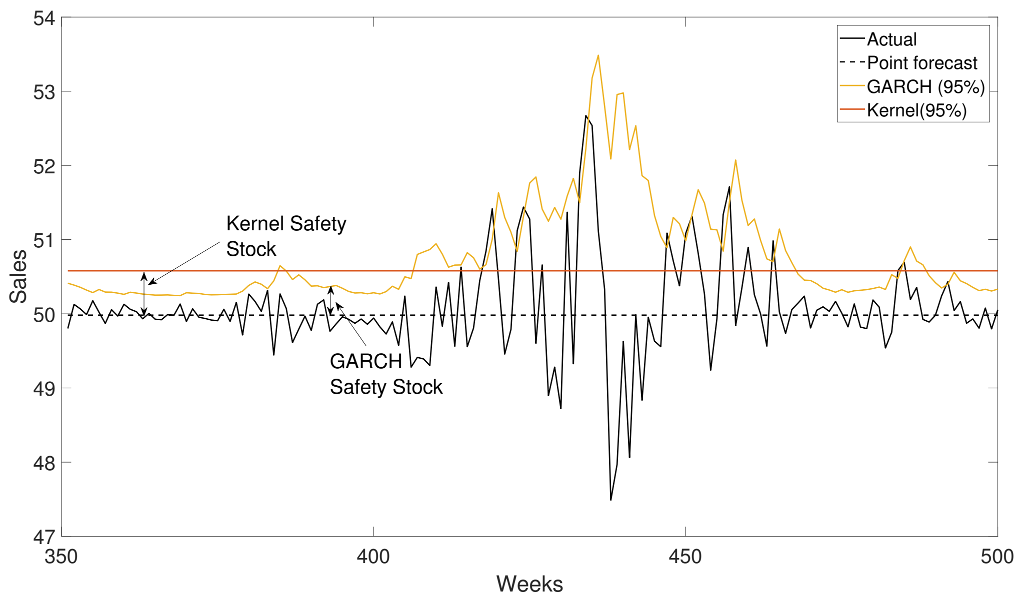

Figure 4: Example of one of the Monte Carlo repetitions

Figure 5 shows that $\sigma_L(4)$ produces a lower CSL and a higher number of backorders than does $\sigma_L(5)$ given the simplifications assumed for $\sigma_L(4)$; thus, our findings agree with those of [20] regarding the inadequacy of $\sigma_L(4)$. The Percentile, Kernel and GARCH methods with $\sigma_L(5)$ achieve CSLs very close to the targets, although the Kernel method does so at a higher scaled safety stock. Interestingly, in terms of the CSL, SES performs worse than the non-parametric approaches, although GARCH performs well. Regarding backorders, GARCH achieves a very good performance—even better than that of $\sigma_L(5)$. In the discussion of the time-varying simulations in the previous section, we concluded that when a parametric model such as GARCH provided lower backorder levels than the rest of the methods it could indicate possible autocorrelation. Figure 5 shows a better performance from GARCH, which possibly indicates that there is some autocorrelation that was not considered in the theoretical Eq. (5) that GARCH can incorporate empirically. We conjecture that such a potential autocorrelation may originate from two sources: i) The first possible source is overlapping temporal demand aggregation, as suggested by the fact that our results agree with those reported by [37], which indicate that the overlapping approach reduces inventory backorder volumes while maintaining the same volumes of held

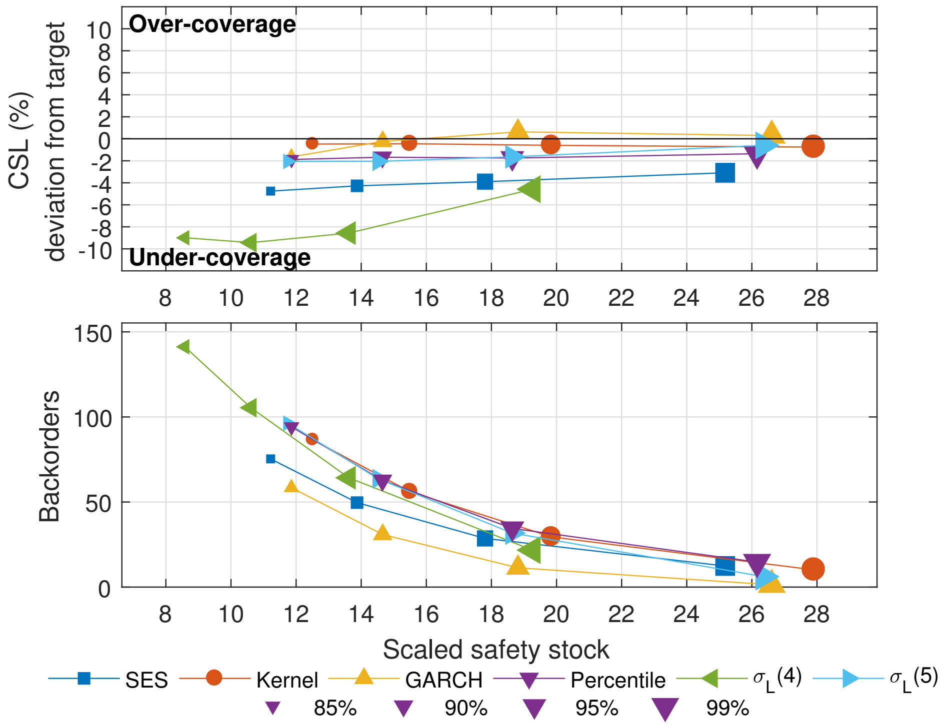

Figure 5: Trade-off curves for a demand that follows an ARIMA(0,1,1) and a lead time of 4 periods

inventory. Note that such variance autocorrelation due to overlapping can be present even when the original demand was independent; ii) Similarly, authors in [10] showed that when the forecasting model parameters are unknown and must be estimated, the forecast errors are correlated even when the demand is independent.

6.4. Stock control policies

In previous simulations, we analyzed the results from the perspective of a newsvendor inventory policy. Then, the question is whether the previous simulations results still hold for different stock control policies. In this section, we answer this question by simulating a periodic review, order-up-to stock control policy, following the work of [51], as shown in Figure 6.

As in the previous simulations, the demand follows a constant level (50 units) plus an ARIMA(0,1,1) with $\theta = -0.75$, a standard deviation of $\sigma = 2$ and a sample size of 500 observations. Each simulation has been repeated

100 times. The ordering decision for an order-up-to policy is as follows:

$$O_t = S_t - IP_t, \tag{16}$$

where O_t, S_t and IP_t denote the ordering decision, order-up-to level and inventory position, respectively, at time t. Note that the review interval is 1. The inventory position can be defined in terms of the net stock (NS_t) and outstanding orders (OO_t):

$$IP_t = NS_t + OO_t, \tag{17}$$

where NS_t is

$$NS_t = \frac{1}{1 - z^{-1}}(R_t - y_t), \tag{18}$$

where Z^{-1} is a Z-transform such as $Z^{-1}y_t = y_{t-1}$ and R_t represents the received orders. We use the variable NS_t to compute the typical stock control metrics. In other words, the achieved CSL is calculated as the proportion of times that the NS_t is greater than zero given that a new replenishment order is launched every time period so that the cycle is unit. Backorders are calculated by summing the negative values of NS_t across time and subsequently averaging across SKUs. Finally, inventory investment is estimated as the average of positive NS_t values across time and SKUs.

The order-up-to level is updated every period:

$$S_t = F_L + k\sigma_L, \tag{19}$$

where F_L is the lead time forecast over L periods. Thus, for every period, the retailer updates the order-up-to level with the current estimates of F_L and σ_L. Note that S_t requires the estimation of the safety stock $k\sigma_L$.

The main difference between the block diagram model shown in [51] and the one shown in Figure 6 is for σ_L. In [51], σ_L is not computed, and the uncertainty is considered as the forecasting demand in one period (i.e., increasing the lead time by one period). In our work, we focus on different ways to compute σ_L and how those affect the safety stock performance. For this simulation, we use Eq. (5) to compute σ_L, although any of the methods presented could be used because we are analyzing the relations between the newsvendor and order-up-to stock control metrics. The described simulation is implemented in SIMULINK. Recall that L already includes a nominal one-period order delay (Z^{-1}) because of the sequence of events [51]; therefore,

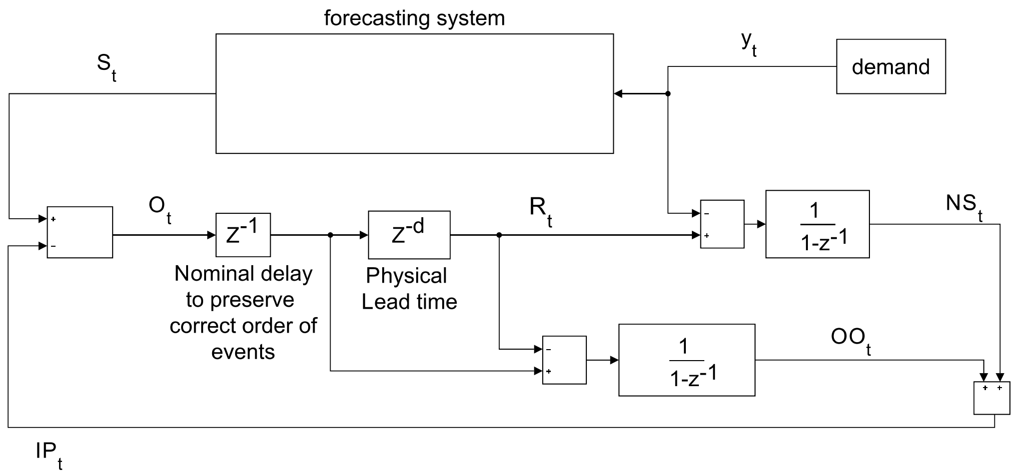

Figure 6: Order-up-to stock control policy plus forecasting system based on Eq. (5)

we now have $L = d + 1$, where Z^{-d} is the physical lead time delay. To be coherent with the previous section, the lead time is set to $L = 4$; therefore $d = 3$. The sequence of events is the following [51]: 1. receive, 2. satisfy demand, 3. count inventory, 4. place order.

Figure 7 shows the relations between the newsvendor performance metrics and the order-up-to stock control metrics. In the three plots, the x and y axes represent the order-up-to metrics and the newsvendor metrics, respectively, for target CSLs of 85%, 90%, 95% and 99%, as in the previous simulations. We observe a high level of correlation between the metrics, as shown by the linear fits in all plots. In other words, the achieved CSL and backorder metrics achieved by the two stock control policies are approximately equal for every target CSL. The main difference lies in the inventory investment of the two techniques; the inventory investment of the newsvendor policy, as measured by the scaled safety stock, is greater than the inventory investment of the order-up-to policy measured by the average net stock. Nevertheless, as in the case of the other metrics, the correlation between the scaled safety stock and the average net stock for different target CSLs is very high. Therefore, the simulation results reported in the previous sections for the newsvendor stock policy can be easily extrapolated to another well-known order-up-to stock control policy.

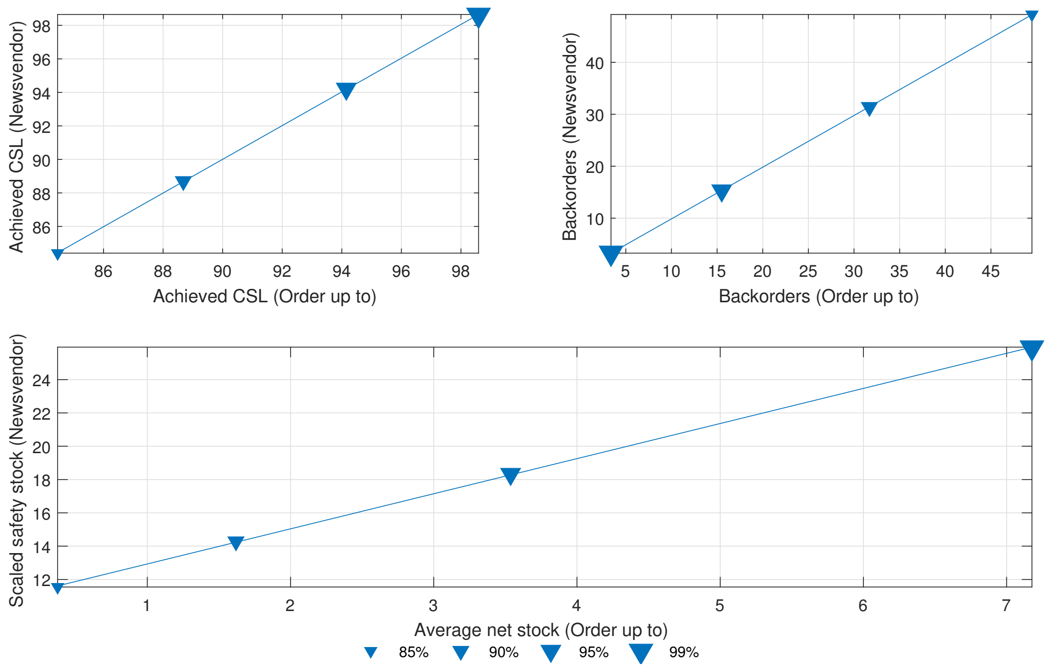

Figure 7: Relationship between newsvendor and stock control policy metrics for target CSL values set to 85%, 90%, 95% and 99%

7. Case study data

The data employed in this paper were previously used by [48]. These data come from a major UK fast-moving consumer goods manufacturer that specializes in the production of household and personal care products. In total, 229 products with 173 weekly observations per product are available. According to [48] the SKUs did not present seasonality, and approximately 21% of them presented trend. We employed SES to produce the point forecasts, although that may not always be the best possible forecasting method. This can be used to assess the robustness of the proposed methods to point forecasts obtained from a method that does not perfectly match with the unknown underlying data-generating process. This is a common situation in both research and in practice given that identifying the true process is not trivial, as mentioned in the Introduction. To verify that SES is a reasonable method to provide point forecasts for our dataset, we compared the forecasting accuracy of the SES with that of ARIMA(p,d,q) models automatically identified by minimizing the Bayesian Information Criterion (BIC) [52] for p=1,2 and 3; q=1,2 and 3; and d=0 and 1. For the point forecasting exercise, we designated 70% of the data as the hold-in sample and the rest as the hold-out sample. We normalized the sales for each SKU with respect the corresponding mean in the hold-in sample to aggregate the results of all SKUS. The results are summarized in Table 1.

Table 1: Comparison of forecasting accuracy between SES and ARIMA models

Error metric	SES	Automatic ARIMA
Mean(RMSE)	0.64	0.72
Median(RMSE)	0.6	0.65

In Table 1, Mean(RMSE) denotes the result of computing the root mean square error for each SKU on the hold-out sample and then averaging across all SKUs. By computing the median instead of the mean of the per-SKU RMSEs, we obtain the Median(RMSE) results reported in the second row. Overall, the SES method achieves a lower forecasting error than the automatic ARIMA models. This is not surprising since, as mentioned previously, the data do not exhibit any marked trends and/or seasonality.

Regarding the volatility forecasting exercise, as in the simulations, the data were split into three subsets. Again, SKU sales were normalized with respect to the hold-in sample means to aggregate the results for all SKUS in the trade-off curves.

To test the normality and variance independence of the forecasting errors, Table 2 shows two statistical metrics that test the residuals normality (the Jarque-Bera test) and conditional heteroscedasticity (the Engle test), both conducted at a 5% significance level. These tests were applied to different lead times that ranged from 1 to 4 weeks. The values of each column represent the percentage of SKUs that do not pass the null hypothesis of normality and no conditional heteroscedasticity. The table was computed using the forecasting errors resulting from removing the training set required to provide the point forecasts. For the minimal lead time of 1 week, 88.2% of the SKUs reject the null hypothesis of normality and 28.2% reject the null hypothesis of no conditional heteroscedasticity. Note the influence of lead time on the statistical tests. As the lead time increases, the percentage of SKUs that do not pass the normality test decreases, as a consequence of the central limit theorem. In contrast, the percentage of SKUs that present heteroscedasticity increases with the lead time. This can be due to the overlapping aggregation process and the estimation of the parameters, as explained in section 6.3, where the influence of different lead times was analyzed.

Table 2: SKU percentages that do not pass the Jarque-Bera and Engle tests at a 5% significance level for different lead times

Lead Time	Jarque-Bera test (%)	Engle test (%)
1	88.2	28.4
2	82.1	91.7
3	74.7	98.7
4	69.4	98.7

To test the adequacy of GARCH(1,1) model on the actual dataset, we have computed the BIC for a general GARCH(p,q) with p=1,2,3 and q=1,2,3 employing the second part of data devoted to optimize the volatility models and a lead time equal to 4. The results show that GARCH(1,1) is a valid model because minimized the BIC in 97% out of total SKUs.

Figures 8 and 9 shows the trade-off curves of the manufacturer data for lead times of 1 and 4 weeks, respectively. In Figure 8, because the lead time is equal to 1, SES, $\sigma_L(4)$ and $\sigma_L(5)$ provide the same results; thus, for the sake of clarity, only SES is plotted. Given that the conditional heteroscedasticity effect is lower for this lead time, the non-parametric Kernel method provides fewer deviations from the target for most CSLs (except 90%). For the highest target CSL (99%), the Kernel method achieves the lowest CSL deviation and the lowest level of backorders, although at the expense of an increased scaled safety stock. This shows evidence of significant skewness in the data and is coherent with the simulation results, where the non-parametric approaches such as Kernel provided better results for non-normal residuals.

Figure 9 considers a lead time of 4 weeks and also shows the theoretical approaches ($\sigma_L(4)$ and $\sigma_L(5)$); $\sigma_L(5)$ obtains a lower CSL deviation and a lower level of backorders than does $\sigma_L(4)$, as expected. Regarding the empirical non-parametric approaches, the Kernel method produces a lower deviation from the target CSL and a lower level of backorders with a higher scaled safety stock. Finally, the empirical parametric approaches (SES and GARCH) show very good performances, especially the GARCH model. Clearly, the GARCH method obtains the lowest CSL deviation among all methods at a similar level of scaled safety stock. Additionally, the GARCH method performs very well with respect to backorders. Note that this better performance of GARCH indicates that the SKUs from this dataset, in general terms, present lead time forecast errors with important variance autocorrelations. This indication is corroborated by the Engle test results in Table 2, and it aligns with the simulation results shown in Figure 3.

8. Conclusions

Despite attention from both academics and practitioners to supply chain risk management, the links between demand uncertainty and risks are still under-researched. One tool that supply chains typically employ to prevent further risks is safety stock. This work examines empirical approaches, both parametric and non-parametric, to estimate the variability of the forecast errors and determine the appropriate safety stocks. In addition, the empirical methods are compared to traditional theoretical methods described in the supply chain literature. Our intention was to provide recommendations for cases in which the assumption of normal i.i.d. forecast errors does not hold, which is the norm in practice.

The results of simulations show that empirical non-parametric approaches such as the Kernel method are suitable when the statistical distribution of the forecast error cannot be assumed to be normal. Additionally, if the variance independence assumption cannot be guaranteed, empirical parametric approaches such as SES and GARCH offer a promising alternative. More-

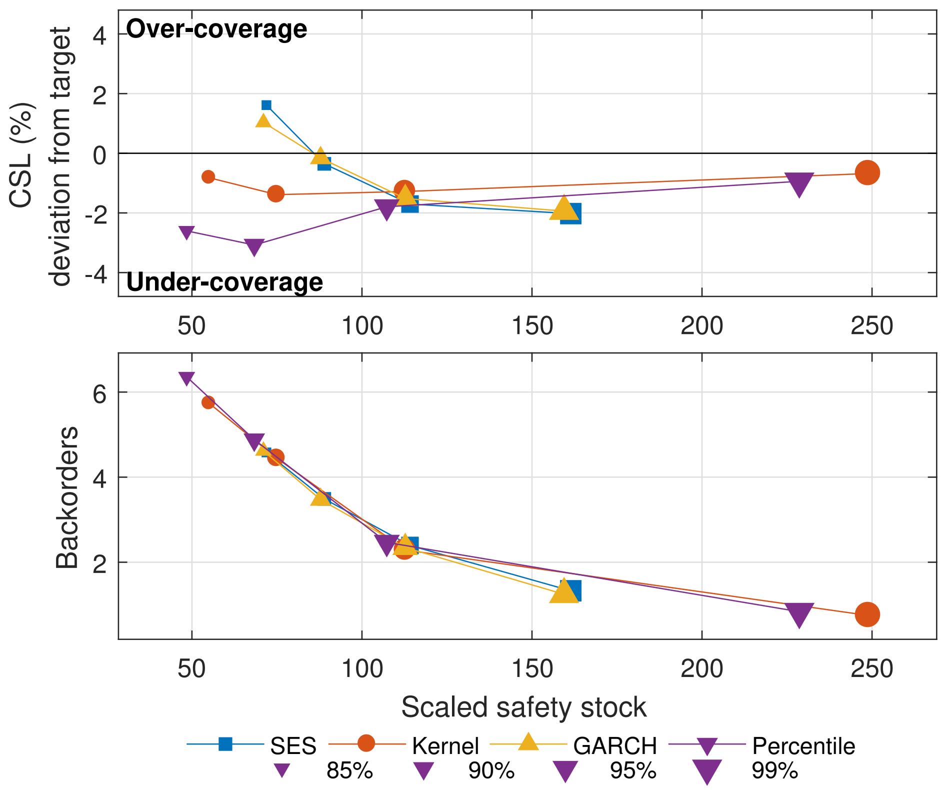

Figure 8: Trade-off curves for the manufacturer data assuming a lead time equal to 1 week

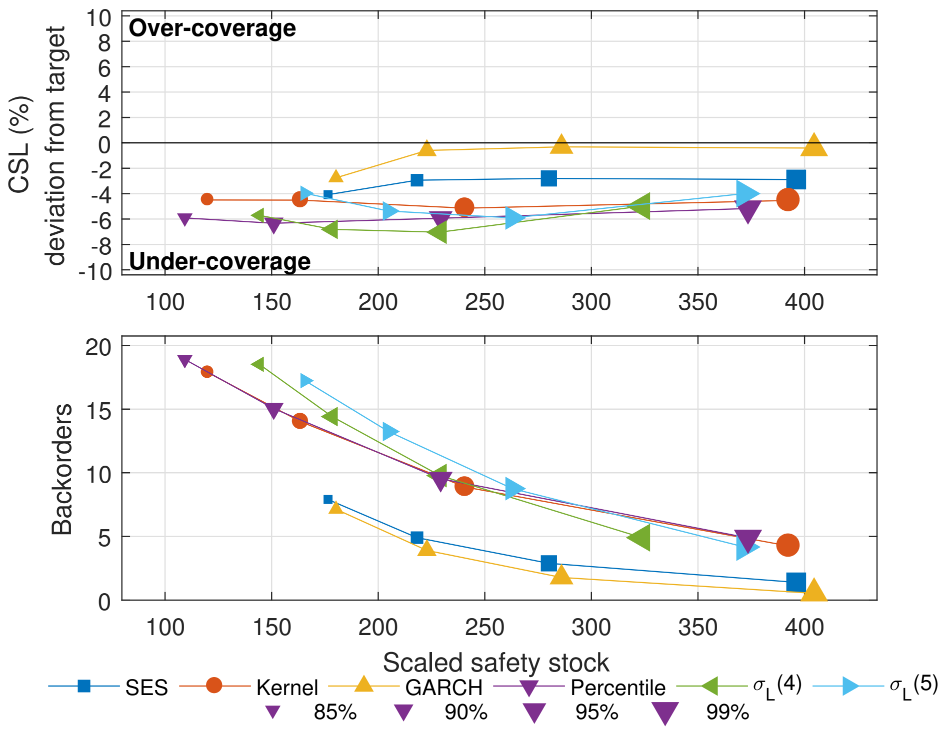

Figure 9: Trade-off curves for the manufacturer data assuming a lead time equal to 4 weeks

over, we find that when GARCH or SES improves the level of backorders, such improvement indicates temporal heterocedasticity on the forecasting errors. Under such circumstances, GARCH is capable of capturing that heterocedasticity and achieves better performance in terms of CSL and backorder volume.

These simulation experiments were assessed for a newsvendor stock policy and were related to the outcome obtained by replacing the newsvendor policy for an order-up-to level. The results showed a high correlation between the inventory performance metrics regardless of the stock control policy.

We also conducted experiments on real data and analyzed the influence of lead time on the i.i.d. assumption. For shorter lead times that were mostly influenced by non-normality, the Kernel method produced better results, particularly for higher CSLs. As the lead time increased, conditional heteroscedasticity became dominant, which caused the empirical parametric methods (GARCH and SES) to provide better results. GARCH outperformed SES in terms of the CSL, because SES is simply a particular case of the general GARCH framework. These results on real data validate the conclusions obtained from the simulation exercises.

The managerial implications of this work are the following. Although the theoretical Eq. (4) is widely used because of its simplicity, if the forecasting model and parameters are known for a certain lead time, we recommend employing a more precise expression, such as Eq. (5), which is valid for the particular case of the simple exponential smoothing forecasting model. Nevertheless, expression (5) does not consider the case in which the forecasting model parameters are not known and must be estimated and nor how the cumulative lead time demand forecasts are computed (overlapping/non-overlapping). For long lead times, these factors can introduce variance autocorrelation and thereby degrade the service level, inventory investment and backorder metrics.

If sufficient data are available and the forecasting model and its parameters cannot be determined (either because the forecasting support system does not provide such information, because the forecasts are obtained using a combination of forecasting methods, or because there are doubts about the i.i.d. assumption with regard to the forecasting errors), then we suggest the use of empirical approaches. For shorter lead times, the Kernel method captures the deviations from normality; for longer lead times, GARCH models are highly suitable and are a good alternative to SES, which has been traditionally utilized since the nineteen-sixties [50]. In this sense, GARCH models can be a good choice, even when the demand variance is uncorrelated.

Future research should address some limitations of this work. Essentially, we analyzed SES and the GARCH(1,1) model for point and volatility forecasts, respectively. Future works should extend GARCH models to incorporate exogenous variables that are available in advance, such as discounted prices due to marketing campaigns. With regard to non-parametric methods, the optimal selection of the kernel function and/or bandwidth also deserves further research. We also anticipate that an analytical contribution that could find a theoretical expression between lead time forecasting error variability resulting from the overlapping aggregation process and the GARCH model will be important.

Acknowledgment

This work was supported by the European Regional Development Fund and Spanish Government (MINECO/FEDER, UE) under project reference DPI2015-64133-R. The authors would also like to thank the anonymous referees for their useful critics and suggestions.

References

[1] B. Fahimnia, C. S. Tang, H. Davarzani, J. Sarkis, Quantitative models for managing supply chain risks: A review, European Journal of Operational Research 247 (1) (2015) 1 – 15.

[2] I. Heckmann, T. Comes, S. Nickel, A critical review on supply chain risk - definition, measure and modeling, Omega 52 (2015) 119 – 132.

[3] E. Silver, D. Pyke, D. Thomas, Inventory and Production Management in Supply Chains. Fourth Edition, CRC Press. Taylor and Francis Group, 2017.

[4] J. Heizer, B. Render, C. Munson, Operations Management: Sustainability and Supply Chain Management: Twelfth edition, Pearson Education, 2017.

[5] S. Nahmias, T. Olsen, Production and Operations Analysis: Seventh Edition, Waveland Press, 2015.

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

[7] E. Porras, R. Dekker, An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods, European Journal of Operational Research 184 (1) (2008) 101 – 132.

[8] Y. S. Lee, A semi-parametric approach for estimating critical fractiles under autocorrelated demand, European Journal of Operational Research 234 (1) (2014) 163 – 173.

[9] A. L. Beutel, S. Minner, Safety stock planning under causal demand forecasting, International Journal of Production Economics 140 (2) (2012) 637 – 645.

[10] D. Prak, R. Teunter, A. Syntetos, On the calculation of safety stocks when demand is forecasted, European Journal of Operational Research 256 (2) (2017) 454 – 461.

[11] D. Prak, R. Teunter, A general method for addressing forecasting uncertainty in inventory models, International Journal of Forecasting. In press.

[12] J. M. Charnes, H. Marmorstein, W. Zinn, Safety stock determination with serially correlated demand in a periodic-review inventory system, Journal of the Operational Research Society 46 (8) (1995) 1006–1013.

[13] T. L. Urban, Reorder level determination with serially-correlated demand, Journal of the Operational Research Society 51 (6) (2000) 762–768.

[14] J. R. Trapero, N. Kourentzes, R. Fildes, Identification of sales forecasting models, J Oper Res Soc 66 (2) (2015) 299–307.

[15] N. Kourentzes, F. Petropoulos, Forecasting with multivariate temporal aggregation: The case of promotional modelling, International Journal of Production Economics 181, Part A (2016) 145 – 153.

[16] X. Zhang, Inventory control under temporal demand heteroscedasticity, European Journal of Operational Research 182 (1) (2007) 127 – 144.

[17] S. Datta, C. W. J. Granger, M. Barari, T. Gibbs, Management of supply chain: an alternative modelling technique for forecasting, Journal of the Operational Research Society 58 (11) (2007) 1459–1469.

[18] S. Datta, D. P. Graham, N. Sagar, P. Doody, R. Slone, O.-P. Hilmola, Forecasting and Risk Analysis in Supply Chain Management: GARCH Proof of Concept, Springer London, London, 2009, pp. 187–203.

[19] Y. Pan, R. Pavur, T. Pohlen, Revisiting the effects of forecasting method selection and information sharing under volatile demand in SCM applications, IEEE Transactions on Engineering Management 63 (4) (2016) 377–389.

[20] C. Chatfield, Time-series forecasting, CRC Press, 2000.

[21] R. Fildes, P. Goodwin, M. Lawrence, K. Nikolopoulos, Effective forecasting and jugdmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning, International Journal of Forecasting 25 (2009) 3–23.

[22] J. R. Trapero, D. J. Pedregal, R. Fildes, N. Kourentzes, Analysis of judgmental adjustments in the presence of promotions, International Journal of Forecasting 29 (2) (2013) 234 – 243.

[23] R. Fildes, Research into Forecasting Practice, Foresight: The International Journal of Applied Forecasting (44) (2017) 39–46.

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

[25] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31 (3) (1986) 307 – 327.

[26] A. A. Syntetos, J. E. Boylan, Demand forecasting adjustments for service-level achievement, IMA Journal of Management Mathematics 19 (2) (2008) 175–192.

[27] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, Bristol, 1986.

[28] O. Isengildina-Massa, S. Irwin, D. L. Good, L. Massa, Empirical confidence intervals for USDA commodity price forecasts, Applied Economics 43 (26) (2011) 3789–3803.

[29] L. Strijbosch, R. Heuts, Modelling (s, Q) inventory systems: Parametric versus non-parametric approximations for the lead time demand distribution, European Journal of Operational Research 63 (1) (1992) 86 – 101.

[30] RiskMetrics, Riskmetrics technical document, Tech. rep., J.P. Morgan/Reuters (1996).

[31] S. Axsäter, Inventory Control: Third edition, International Series in Operations Research & Management Science, Springer International Publishing, 2015.

[32] W. E. Wecker, The variance of cumulative demand forecasts, in: Graduate School of Business, Univ. of Chicago Waterloo, Canada Working Paper, 1979.

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

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

[35] R. J. Hyndman, A. B. Koehler, J. K. Ord, R. D. Snyder, Forecasting with Exponential Smoothing: The State Space Approach, Springer-Verlag, Berlin, 2008.

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

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

[38] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, Econometrica: Journal of the Econometric Society 50 (1982) 987–1007.

[39] D. B. Nelson, Stationarity and persistence in the GARCH(1,1) model, Econometric Theory 6 (3) (1990) 318–334.

[40] E. S. Gardner, Exponential smooothing: The state of the art, Journal of Forecasting 4 (1) (1985) 1–28.

[41] E. S. Gardner, Exponential smoothing: The state of the art, Part II, International Journal of Forecasting 22 (2006) 637–666.

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

[43] N. Kourentzes, On intermittent demand model optimisation and selection, International Journal of Production Economics 156 (2014) 180 – 190.

[44] A. A. Syntetos, M. Z. Babai, E. S. Gardner, Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping, Journal of Business Research 68 (8) (2015) 1746 – 1752.

[45] G. Cachon, C. Terwiesch, Matching supply with demand: an introduction to operations management: Third edition, McGraw-Hill Education, 2013.

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

[47] L. Tashman, Out-of-sample tests of forecasting accuracy: an analysis and review, International Journal of Forecasting 16 (2000) 437–450.

[48] D. Barrow, N. Kourentzes, Distributions of forecasting errors of forecast combinations: implications for inventory management, International Journal of Production Economics 177 (2016) 24–33.

[49] T. A. Burgin, The gamma distribution and inventory control, Journal of the Operational Research Society 26 (3) (1975) 507–525.

[50] R. Brown, Statistical forecasting for inventory control, New York: McGraw-Hill, 1959.

[51] J. Dejonckheere, S. M. Disney, M. R. Lambrecht, D. R. Towill, Measuring and avoiding the bullwhip effect: A control theoretic approach, European Journal of Operational Research 147 (2003) 567–590.

[52] G. Schwartz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.