

Please cite this paper as:

Trapero J. R., Cardos M., Kourentzes N., 2016. Empirical safety stock estimation based on Kernel and GARCH models. Management Science Working Paper Series, Lancaster University, Department of Management Science.



Lancaster University
Management School

Lancaster University Management School
Working Paper 2016:5

Empirical safety stock estimation based on Kernel and GARCH models

Juan R. Trapero, Manuel Cardos and Nikolaos Kourentzes

*The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK*

© Juan R. Trapero, Manuel Cardos and Nikolaos Kourentzes
All rights reserved. Short sections of text, not to exceed
two paragraphs, may be quoted without explicit permission,
provided that full acknowledgment is given.

The LUMS Working Papers series can be accessed at <http://www.lums.lancs.ac.uk/publications>
LUMS home page: <http://www.lums.lancs.ac.uk>

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 4Y X, UK

Abstract

Supply chain risk management is drawing the attention of practitioners and academics. A source of risk is demand uncertainty. To deal with it demand forecasting and safety stocks are employed. Most of the work has focused on point demand forecasting, assuming that forecast errors follow the typical normal i.i.d. assumption. The variability of the forecast errors is used to compute the safety stock, in order to reduce the risk of stockouts with a reasonable inventory investment. Nevertheless, real products' demand is very hard to forecast and that means that at minimum the normally i.i.d. assumption should be questioned. This work analyses the effects of possible deviations from these assumptions and it proposes empirical methods based on Kernel density estimators (non-parametric) and GARCH models (parametric) in order to compute the safety stock. The results show that Kernel density estimator is recommended when the forecast errors are fat tailed and GARCH models are well-suited when forecast errors present autocorrelation. Additionally, GARCH models present important improvements for lead time forecast errors, as shown in terms of customer service level, inventory investment and backorders volume. Simulations and real demand data from a manufacturer are used to illustrate our methodology.

*Corresponding author.

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

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 with a growing number of related publications (Fahimnia et al., 2015). However, among the different related works, a clear meaning for SCRM had not emerged. Recently, in order to bridge that gap, Heckmann et al. (2015) accomplished a critical review of SCRM and provided a definition of supply chain risk: “Supply chain risk is the potential loss for a supply chain in terms of its target values of efficiency and effectiveness evoked by uncertain developments of supply chain characteristics whose changes were caused by the occurrence of triggering-events”

Following that definition, one of the supply chain characteristics that is subject to uncertainty is future demand. In this sense, if demand is unusually large, a stockout may occur with the associated negative consequences. If demand is lower than expected, then the company may carry higher holding costs due to excess inventories. In order to mitigate such risks, safety stocks may be employed, which are additional units over the stock required to attend the lead time forecasted demand. According to Silver et al. (1998), different approaches can be utilized to calculate the safety stock: i) safety stocks established through the use of a simple-minded approach; ii) safety stocks based on minimizing cost; iii) safety stocks based on customer service; and, finally, iv) safety stocks based on aggregate considerations.

Although the most appropriate method depends on the organization circumstances, calculating the safety stock based on customer service is widely used given that it does not require knowledge of the stockout cost that can be very difficult to estimate. The safety stock based on customer service is determined on the basis of the demand uncertainty.

A literature survey may suggest to use demand variability as a variable of demand uncertainty. For example, Heizer and Render (2004) suggest this approach. Other authors like Silver et al. (1998) and Nahmias (2009) suggest the use of demand forecasting error variability instead, given that future demand is unknown and it should be forecasted; thus safety stocks are designed to prevent issues due to such demand forecast errors. Additionally, the use of demand variability may under/over-estimate demand uncertainty. Nahmias

(2009) shows that demand variability underestimates demand uncertainty, i.e. the demand forecast error variance is greater than the demand variance when simple exponential smoothing is used as forecasting technique, unless the exponential smoothing constant is equal to zero. When this is true, as long as the initial value for the level is optimal, the forecast is simply the arithmetic mean of the historical sales, and in that case both variances are the same, as the variance formula becomes identical. It should be pointed out that simple exponential smoothing is appropriate for non-trended and non-seasonal demand data generating processes (Gardner, 2006), and therefore the difference in variances is not due to using an inappropriate forecast.

Conversely, in situations when the demand data generator process cannot be assumed to be a constant plus an error term, for example when the product is seasonal or when it is subject to any kind of promotions, the use of demand variance instead of demand forecast error variance may overestimate the demand uncertainty. In the seasonal case, part of the demand variability can be anticipated by using an adequate forecasting model like Holt-Winters exponential smoothing. In the other case, promotions are designed beforehand and causal promotional models can be used (Trapero et al., 2015; Beutel and Minner, 2012). Therefore, in this work, demand forecast errors will be used as the demand uncertainty variable. More details about the difference between variance of demands and variance of forecasts can be found in Katsikopoulos and Syntetos (2016).

Focusing on the demand forecast error, it is assumed to be Gaussian independent and identically distributed (i.i.d.) with zero mean and constant variance. Nonetheless, often the forecast error does not fulfill these assumptions and this may cause a lower service level than targeted. In other words, the risk of a stockout is increased. It should be noted that, although empirical approaches in different areas such as prediction intervals theory and the financial risk management literature have been developed to overcome such limitations, their application to safety stock calculation and, therefore, to supply chain risk management, have remained overlooked.

The aim of this work is to compute the safety stock for a certain lead time by using empirical variability estimations of the measured forecast errors, where the i.i.d. assumption is relaxed. First, we will explore the use of empirical non-parametric approaches, such as the kernel density estimator (Silverman, 1986), which are typically utilized in prediction interval case studies (Isengildina-Massa et al., 2011), in order to avoid the normality assumption on the forecasting error. Second, empirical parametric estimators

associated to financial risk management, such as Generalised Autoregressive Conditional Heteroskedastic (GARCH, Bollerslev, 1986), along with exponential smoothing models (Taylor, 2004) will be investigated to overcome the independence assumption and exploit potential forecast error correlations.

The performance of the aforementioned empirical approaches will be measured first by using prediction intervals metrics, like forecast coverage and average interval width. Subsequently, simulation experiments that incorporate an order-up-to level stock control policy will be used to analyze the close links between prediction interval metrics and classic stock control metrics like cycle customer service level, inventory investment and backorders. The proposed empirical approaches will be compared to the traditional supply chain theoretical approaches based on the i.i.d. assumptions on simulated and real data from a manufacturing company.

The rest of the paper is organised as follows: Section 2 reviews the traditional literature on safety stock calculation that is denoted by theoretical approach. Section 3 describes the empirical approximation proposed in this work. Section 4 explains the performance metrics that will be used to assess the different methods and provides any implementation details of the alternative approaches. Simulation experiments are carried out in Section 5. A case study based on a manufacturer shipments data is described in Section 6, and finally, section 7 presents the concluding remarks.

2. Literature review

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

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

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

The main problem in (1) is how to estimate σ_L . In order to do that, we have two approaches: a theoretical approach, where an estimation of σ_1 (one-step ahead standard deviation of the forecast error) is provided and an analytic expression that relates σ_L and σ_1 is employed. The estimate of

σ_1 is possible as the updating forecast step is usually smaller than the lead time, On the other hand, an empirical approach can be employed, where σ_L is directly estimated from the lead time forecast error. We discuss the theoretical approach in detail below, and the empirical approach in the next section.

2.1. Theoretical approach

2.1.1. Estimation of σ_1

Traditional textbooks suggest to compute σ_1 based on forecast error metrics. For instance, Silver et al. (1998) uses MSE (Mean Squared Error) and Nahmias (2009) suggests MAD (Mean Absolute Deviation):

$$MSE_t = \frac{1}{n} \sum_{t=1}^n (y_t - F_t)^2, \quad (2)$$

$$MAD_t = \frac{1}{n} \sum_{t=1}^n |y_t - F_t|, \quad (3)$$

where y_t is the actual value at time t and F_t is the forecast value for the same time period and n is the sample size. The one step ahead standard deviation of the forecasting error can be approximated by

$$\sigma_1 = \sqrt{MSE_{t+1}}. \quad (4)$$

Note that if we want to use the MAD metric to estimate the standard deviation, under the normal distribution assumption it holds that

$$\sigma_1 = 1.25MAD_{t+1}. \quad (5)$$

As the conversion factor from MAD to $\hat{\sigma}_1$ is different for other statistical distributions and for MAD the results of the analysis would be proportional, hereafter we will simply use MSE to estimate σ_1 . Additionally, the estimation of MSE can be updated when new observations are available such as:

$$MSE_{t+1} = \alpha'(y_t - F_t)^2 + (1 - \alpha')MSE_t, \quad (6)$$

where MSE_{t+1} is the forecast of MSE for the period $t+1$ and α' is a smoothing constant that varies between 0 and 1, although small values between 0.01 and 0.1 are commonly employed (Silver et al., 1998). In this work α' and the initial value are optimized by minimizing the in-sample squared error

following the suggestion by RiskMetrics (1996). Note that (6) is the well known single exponential smoothing.

2.1.2. Estimation of σ_L

A theoretical formula for σ_L can be obtained on the basis of σ_1 , the forecast updating procedure and the values of the smoothing constants used. According to Silver et al. (1998), an exact relationship can be complicated and they propose the following approximation:

$$\sigma_L = L^c \sigma_1, \tag{7}$$

where c is a coefficient that must be estimated empirically, considering different SKUs. It should be noted that in this work we assume the lead time (L) constant and deterministic and thus focus our analysis on σ .

Axsäter (2007) argues that if we disregard that errors usually increase for longer forecast horizons and we also assume that forecast errors are independent over time, then:

$$\sigma_L = \sqrt{L} \sigma_1. \tag{8}$$

It should be noted that when $c = 0.5$ then (7) corresponds the assumption of forecast errors being independent over time, while in the other extreme, $c = 1$ corresponds to forecast error that are completely correlated over time. He also proposes to use $c > 0.5$ to compensate for disregarding that the forecast error for a more distant period tends to be larger.

Hyndman et al. (2008) shows that, in case the demand can be modeled as a local level model, i.e., a ETS(A,N,N) with parameter α , the conditional variance for the lead-time demand is

$$\sigma_L = \sigma_1 \sqrt{L} \sqrt{1 + \alpha(L - 1) + \frac{1}{6} \alpha^2 (L - 1)(2L - 1)}. \tag{9}$$

Note that the difference between (8) and (9) increases as both the lead time and the exponential smoothing constant become larger. The lead time variance for some well-known exponential smoothing models can be found in page 91 of Hyndman et al. (2008). It is interesting to note that although an exact relationship has been provided in (9), most of the literature still is using the approximation given in (8). Furthermore, expression (9) was initially derived by Wecker (1979) for a deterministic lead time for an IMA(0,1,1) demand process, which is equivalent to ETS(A,N,N). According to Graves

(1999a), Wecker’s manuscript was not published but was referenced by Eppen and Martin (1988). On the other hand, Graves (1999b) arrived at the same result for an inventory system that is subject to an IMA(0,1,1) demand process, a deterministic replenishment lead-time and an adaptive base-stock policy.

3. Empirical estimation of the lead time forecast error variability

Note that the theoretical approach assumes that forecast errors are i.i.d. and this 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 Chatfield (2000). Given the complex nature of markets, clients, promotions, economic situation, etc., assuming that we know the true demand model for each SKU would be far from realistic and empirical approaches must be, at least, tested. An empirical estimate of σ_L can be calculated as:

$$\sigma_L = \sqrt{\frac{\sum_{t=1}^n (e_t(L) - \bar{e}(L))^2}{n}}, \quad (10)$$

where $e_t(L) = \sum_{h=1}^L y_{t+h} - \sum_{h=1}^L F_{t+h}$ is the lead time forecast error and $\bar{e}(L)$ is the average error for the L under consideration. In this approach, we do not need to know neither the forecasting model/method nor its parameters. This could be a potential advantage for Forecasting Support Systems that do not provide such information to users. Furthermore, an estimate of σ_1 is not needed.

It should be noted that, although empirical approaches have been shown to give good results in other applications like prediction intervals theory (Chatfield, 2000), its use for supply chain risk management has been scarce. In order to bridge that gap, this section is devoted to show how empirical lead time errors can be used effectively to determine the safety stock, or in other words, to measure the demand uncertainty risk.

In order to employ an empirical approach we have to make two decisions. First, what kind of empirical data is used? This can be in-sample or out-of-sample. Second, whether the method is parametric or not (Lee and Scholtes, 2014).

For the in-sample errors case, the forecasting model is applied to the same data employed for estimating its parameters to provide a sample of forecast

errors at different lead times. However, in-sample errors may also be inadequate given that residuals errors usually are lower than out-of-sample errors (Tashman, 2000; Barrow and Kourentzes, 2016). In order to overcome this limitation, empirical approaches using out-of-sample errors have also been developed, i.e., forecast errors obtained from the forecasting model applied on data that have not been used in the fitting process. This kind of empirical approach was initially suggested by Williams and Goodman (1971) within a prediction interval context and it has gained popularity in diverse applications, particularly in those areas where data requirements are not a limitation (Isengildina-Massa et al., 2011).

In this work we estimate σ_L using out-of-sample forecast errors. To do this, we can employ the following parametric and non-parametric approaches.

3.1. Parametric approach. SES and ARCH/GARCH models

In the empirical parametric approach, we keep the assumption that lead time forecast errors are normally distributed, but we relax the independence assumption by allowing time-varying σ_L .

3.1.1. Single Exponential Smoothing, SES

Expression (6) shows how σ_1 can be updated with the new observed forecast errors. Likewise, we can update σ_L on the basis of the lead time forecast error instead of the one-step-ahead forecasting error:

$$MSE_{L,t+1} = \alpha''(y_L - F_L)^2 + (1 - \alpha'')MSE_{L,t}, \quad (11)$$

where $\sqrt{MSE_{L,t+1}}$ forecasts σ_L in (18) at time $t + 1$. Unlike expression (6), the updating step does not match with the lead time forecast error. The expression is updated at every period, but by using y_L and F_L that correspond to L periods. For example, assume weekly observations and the lead time could be 4 weeks. When a new observation is received both y_L and F_L are recalculated by dropping the oldest observation and adding the new one in the sum over 4. Effectively a period is used for the calculation of y_L and F_L four times. On the other hand, there is a unique $MSE_{L,t+1}$ for each period, since this is updated at every step.

The process followed for the calculation of y_L and F_L is known as overlapping temporal demand aggregation (Boylan and Babai, 2016). In the same reference, the authors recommend the overlapping approach when the demand history is not too short. For instance, they show that the overlapping

approach produce lower variance estimates for a demand history of 24 periods. As we will see later on, our case study has a demand history of 173 periods per SKU and so, the overlapping approach is adopted. Similarly to (6), α'' and the initial value are optimized by minimizing squared errors.

It should be pointed out that by using expressions (6) or (11), we are implicitly assuming that the forecast errors can be dependent and/or heteroskedastic. Otherwise computing the in-sample standard deviation of the forecast errors would be enough given that it remains constant across time. However, from the authors' point of view, the supply chain literature has not achieved important advances regarding how to deal with such a potential heteroskedasticity/autocorrelation of forecast errors. In turn, other disciplines related to financial risk management have done an important effort to deal with this problem providing sophisticated models. In particular, they have widely employed the family of ARCH/GARCH models.

3.1.2. 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 and they may be good candidates for enhancing the risk estimation and therefore, the determination of the safety stock.

Particularly, the AutoRegressive Conditional Heteroscedasticity (ARCH) model introduced by Engle (1982) is one of the most important developments in risk modelling that can be well-suited for our application. Basically, this model expresses the forecast error as $\epsilon_t = \sigma_t \cdot v_t$, where

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^p a_i \epsilon_{t-i+1}^2 \quad (12)$$

and v_t is an independent and identically distributed (i.i.d.) process. ARCH models express the conditional variance as a linear function of p lagged squared error terms. Bollerslev (1986) proposed the generalized autoregressive conditional heteroscedasticity (GARCH) models that represent a more parsimonious and less restrictive version of the ARCH(p) models. Generally speaking, GARCH models with respect to ARCH models bring similar advantages as exponential smoothing with regards to moving average (Boudoukh et al., 1997). 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,

GARCH(1,1) model is given by:

$$\sigma_{t+1}^2 = \omega + a_1 \epsilon_t^2 + \beta_1 \sigma_t^2. \quad (13)$$

It should be noted that exponential smoothing has the same formulation as the integrated GARCH model (IGARCH) (Nelson, 1990), in which $\beta_1 = 1 - a_1$, plus the additional restriction $\omega = 0$. If we apply the GARCH(1,1) on the lead time forecast error instead the one-step ahead forecasting error, equation (13) can be rewritten as:

$$\sigma_{L,t+1}^2 = \omega' + a_1' \epsilon_{L,t}^2 + \beta_1' \sigma_{L,t}^2. \quad (14)$$

In this work, we focus our analysis on the GARCH(1,1) model using an overlapping approach in order to compare it with the Single Exponential Smoothing model in (11). It should be noted that the model GARCH(1,1) also includes the model GARCH(0,1) when the estimation of a_1' is equal to zero.

3.2. Non-parametric approach

It is likely that 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 the forecast errors may not hold. In order to overcome this, non-parametric approaches can be well-suited. If we want to consider non-parametric methods, safety stocks calculation should be reformulated as:

$$SS = Q_L(CSL), \quad (15)$$

where $Q_L(CSL)$ is the forecast error quantile at the probability defined by CSL. Such a quantile can be estimated non-parametrically by means of the empirical distribution of the generated lead time forecast errors (e_L). Here, we are going to use two well-known non-parametric methods: i) Kernel Density Estimator; and ii) Empirical Percentile.

3.2.1. Kernel density estimator

This technique represents the probability density function ($f(x)$) of the lead time forecast errors without the need of making assumptions about the distribution of the data. Its formula for a series X at a point x is given by:

$$f(x) = \frac{1}{Nh} \sum_{j=1}^N K\left(\frac{x - X_j}{h}\right), \quad (16)$$

where N is the sample size, $K(\cdot)$ is the kernel smoothing function that integrates to one and h is the bandwidth (Silverman, 1986). For instance, the CSL quantile denoted by $Q_L(CSL)$ can be estimated non-parametrically by means of the empirical distribution of the generated lead time forecast errors.

3.2.2. Empirical percentile

The well known percentile indicates the value below which a given percentage of observations fall. For example, the 90th percentile is the value under which 90% of the sampled observations can be found. Typically in the calculation of empirical percentiles the requested value is linearly interpolated from the available ones. For instance if there are 5 values available, when sorted these will correspond to the 10th, 30th, 50th, 70th and 90th percentiles. The 80th percentile will be calculated using a linear interpolation between the last two values.

4. Evaluation of the alternative approaches

In this section we present the various metrics that are used to evaluate the performance of the approaches described above. In addition to the typical inventory metrics, in this work, we also proposed the use of prediction interval metrics. Furthermore, we provide in this section any relevant implementation details. The specifics of the various datasets used in the evaluation are outlined in the simulation and case study results respectively.

4.1. Point forecast

Before assessing the different alternatives to estimate the forecast error variability, it is required to provide point forecasts. In this work the single exponential smoothing will be used for two reasons, firstly, given its wide use in business applications (Gardner, 1985, 2006) and secondly, for comparison purposes since theoretical formula in (9) is based on a simple exponential smoothing, that can be expressed as:

$$F_{t+h} = \alpha y_t + (1 - \alpha)F_t, \quad (17)$$

where $0 < \alpha < 1$. Given the recursive nature of exponential smoothing, it is necessary to initialize the algorithm. We optimize the initial value together with the α value by minimizing the in-sample mean squared error. Note that the lead time forecast is $F_L = L \cdot F_{t+1}$. In order to have a fair comparison between variance forecasting methods we will use this point forecast in all cases.

4.2. Prediction interval metrics

The safety stock calculation problem can be seen as the problem of finding a prediction interval such as:

$$[L_L, U_L] = [0, F_L + k \cdot \sigma_L], \quad (18)$$

where $F_L = \sum_{h=1}^L F_{t+h}$ and F_{t+h} is the point forecast at $t+h$, L_L is the lower interval set to zero and U_L the upper interval that is the sum of the demand forecast plus the safety stock for a determined lead time L . Recall that the safety stock is $SS = k \cdot \sigma_L$. Moreover, to determine k we will use one-sided distribution because we are interested in the case where the demand is higher than the upper limit, because it would result in a stockout.

Generally speaking, prediction intervals are assessed on the basis of two metrics. Firstly, the forecast coverage, and secondly, the average prediction interval width. The former measures the proportion of times that a real value falls inside the prediction interval given a certain target confidence level. The latter evaluates how wide or narrow is the provided prediction interval. The ideal situation is a method that provides a forecast coverage close to the target level with a low interval width.

4.3. Inventory performance metrics

In a supply chain context, according to Gardner (1990), Kourentzes (2014) and Syntetos et al. (2015), trade-off curves are employed to measure the forecasting performance of different techniques. These trade-off curves can be the couples formed by customer service level and inventory investment plus backorders and inventory investment. Kourentzes (2014) transformed inventory investment and backorders metrics into scale independent by dividing them with the average in-sample demand, to allow easy summarizing of the metrics.

4.4. Links between prediction intervals and inventory performance metrics

Interestingly, the forecast coverage and interval width from prediction intervals theory provide similar information to customer service level and inventory investment, respectively, from stock control performance metrics. In other words, if our prediction interval provides a good forecast coverage, we expect that the customer service level is going to be good as well. Likewise, if the prediction interval width is narrow, we expect that the inventory investment were low too.

It should be remarked that there is an important difference, in order to use stock control metrics we have to define a stock control policy, whereas prediction interval metrics are independent of such stock control policies.

In order to investigate the connection between prediction intervals and inventory performance metrics, by the end of the next section, we are going to simulate a typical order-up-to stock control policy and compare the results obtained by the prediction interval metrics (forecast coverage, prediction interval width) and those provided by inventory performance metrics (customer service level, inventory investment). In fact, since we are interested in the safety stock, the average prediction interval width is going to be measured by the so-called scaled safety stock, which is defined as the prediction interval at time t minus the point forecast at the same time, all divided by the in-sample demand average. Additionally, we will also include the variable backorders calculated in two steps: 1) the sum of units out of the prediction interval per each SKU on the hold out sample; and 2) the average of that sum across SKUs.

In this work the target confidence level and target customer service level are considered equivalent and set to: 85%, 90%, 95% and 99% for both simulations and real data.

4.5. Implementation of approaches

The analysis is carried out in MATLAB. Here we briefly list the implementation details of the approaches discussed above. The financial toolbox from MATLAB was employed to implement the GARCH(1,1) model. We have used Kernel density estimators from MATLAB with its default options for the selection of the bandwidth, and Epanechnikov kernel. The empirical percentile computed using the function “prctile” from MATLAB.

5. Simulation results

In this section we carry out four simulation experiments 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 homoskedastic and independence assumptions do not hold; and (iii) how different lead times affect at those results. Those simulations will be assessed by means of prediction interval metrics (forecast coverage, scaled safety stock and backorders). In order to connect these results to a supply chain context, the last simulation (iv) will analyse the relationship between prediction interval metrics and inventory performance metrics under an order-up-to level stock policy.

In these experiments we have divided the total sample in three parts. The first 40% of the data is used to estimate the initial value and exponential smoothing parameter to provide the mean forecasts. The following 40% of the data is used to estimate the initial value and exponential smoothing parameter to produce volatility forecasts, as well as, the Kernel density estimate, empirical percentile and GARCH(1,1) parameters. The rest of the data is kept as hold-out sample for evaluation purposes.

5.1. Sample size and distribution shape

In order to explore the influence of demand distribution on the safety stock calculation we will employ two distributions: (i) normal; and (ii) log-normal. This permits us to observe any differences found if the demand is skewed. Note that the lognormal distribution is reasonable to appear when products are subject to promotional periods, during which the observed demand is higher than the baseline demand.

The sample size is an important variable to study, given that often products have short life cycles and there are not many historical observations available. We run simulations for the following sample sizes: 50, 100 and 500. For each sample size and each distribution (normal and lognormal) we have carried out 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 $\sigma = 25$. Regarding the log-normal distribution, the parameter values were $\mu = 3$ and $\sigma = 1$. Note that the log-normal distribution does not yield negative values.

Figure 1 shows the trade-off curves for a sample size of 50 observations and a lead time equal to 1. The upper plots show the coverage deviation from the target in percentage. For example, for a target at 85%, denoted with the smallest marker, the coverage deviation for SES is about -1%, i.e., it obtained a coverage of 84%. The closer the lines are to zero, the better. Lower panels plot the relationship between backorders and scaled safety stock. Note that the different target coverages are organized from the smallest size marker (85%) to the largest size (99%).

We do not show the results for the GARCH(1,1) model in Figure 1 as there were numerical problems with its estimation. Since the simulated errors are independent, i.e. the forecast errors do not have any structure, the GARCH model is not adequate. The left side of that figure (upper and lower panels) shows the trade-off curves for the normally distributed demand. As we expected, SES performs reasonably well providing a systematic under-coverage, whereas Kernel provides a systematic over-coverage that is reduced for the highest target coverage. In this case, the empirical percentile achieves a remarkable lower coverage when the target is (99%), that also implies a higher backorders volume.

On the same figure right side, we find the same trade-off curves when the simulated demands are log-normal. Essentially, non-parametric approaches (Kernel and Percentile) achieve the highest and lowest extreme targeted confidence levels (99% and 85%), whereas parametric approach (SES) did not. Therefore, in case the forecast errors are skewed the typical normal assumption may induce lower service levels for a high confidence target (99%) and an overcoverage for low confidence target (85%). In this sense, non-parametric approaches seem more robust.

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 service levels (99% and 85%) for demands log-normally distributed. When the distribution is normal, all the methods achieve a similar performance very close to the target confidence levels, although the Kernel provides systematically coverage levels higher than the expected.

Figure 3 shows the same simulations if we increase the sample size to 500 observations. Again, non-parametric approaches achieve a better performance for log-normal demand distribution, whereas the parametric approach does not achieve the extreme service levels (85 % and 99 %). When the demand is normal all the methods show a similar performance, although the

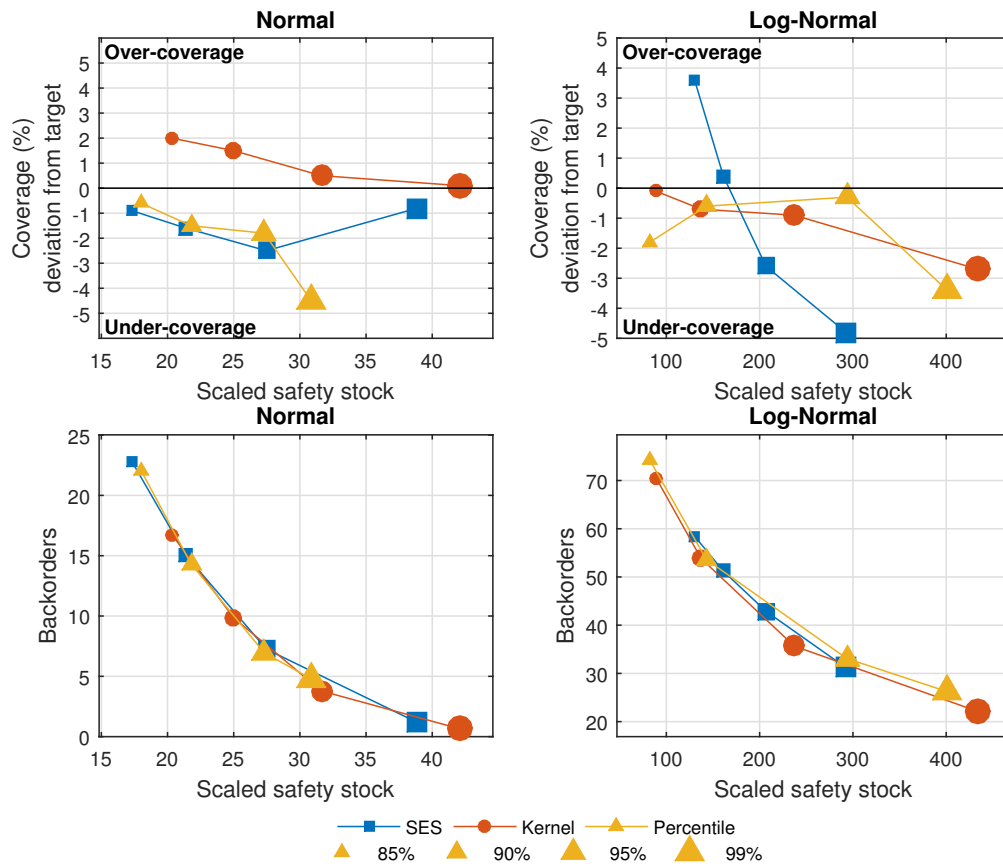


Figure 1: Trade-off curves for a sample size of 50 observations. Upper and lower left panels: Normal distribution. Upper and lower right panels: Log-normal distribution.

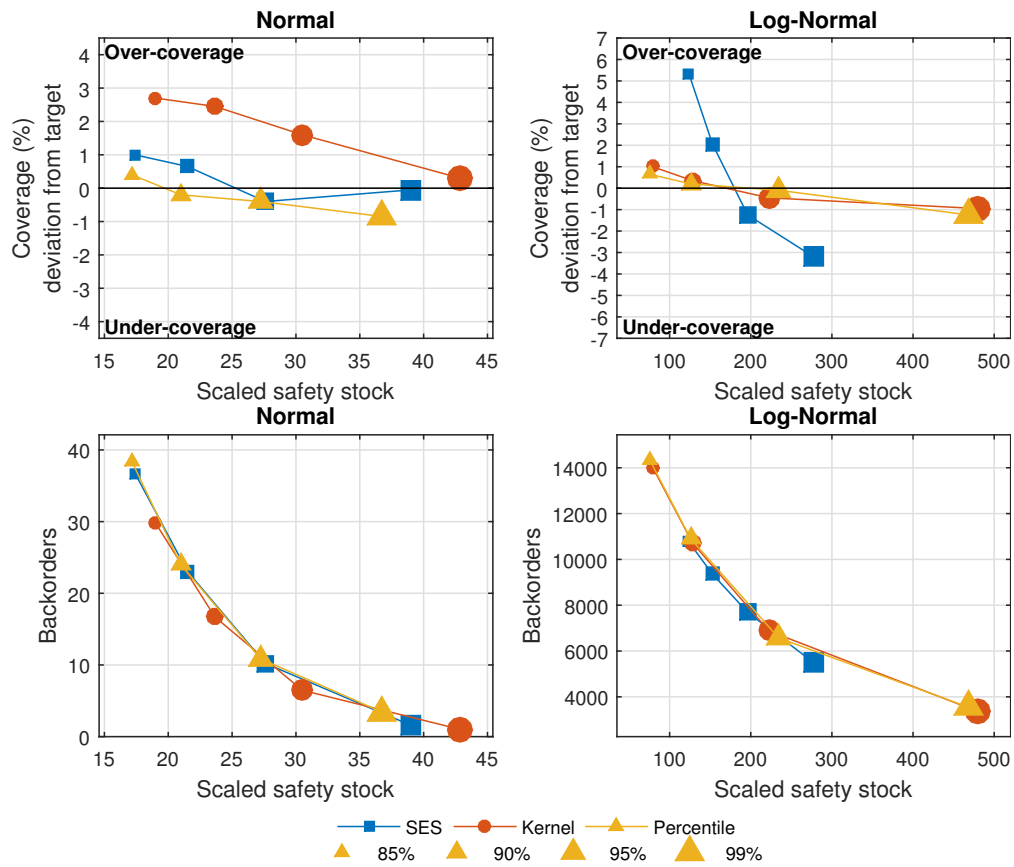


Figure 2: Trade-off curves for a sample size of 100 observations. Upper and lower left panels: Normal distribution. Upper and lower right panels: Log-normal distribution.

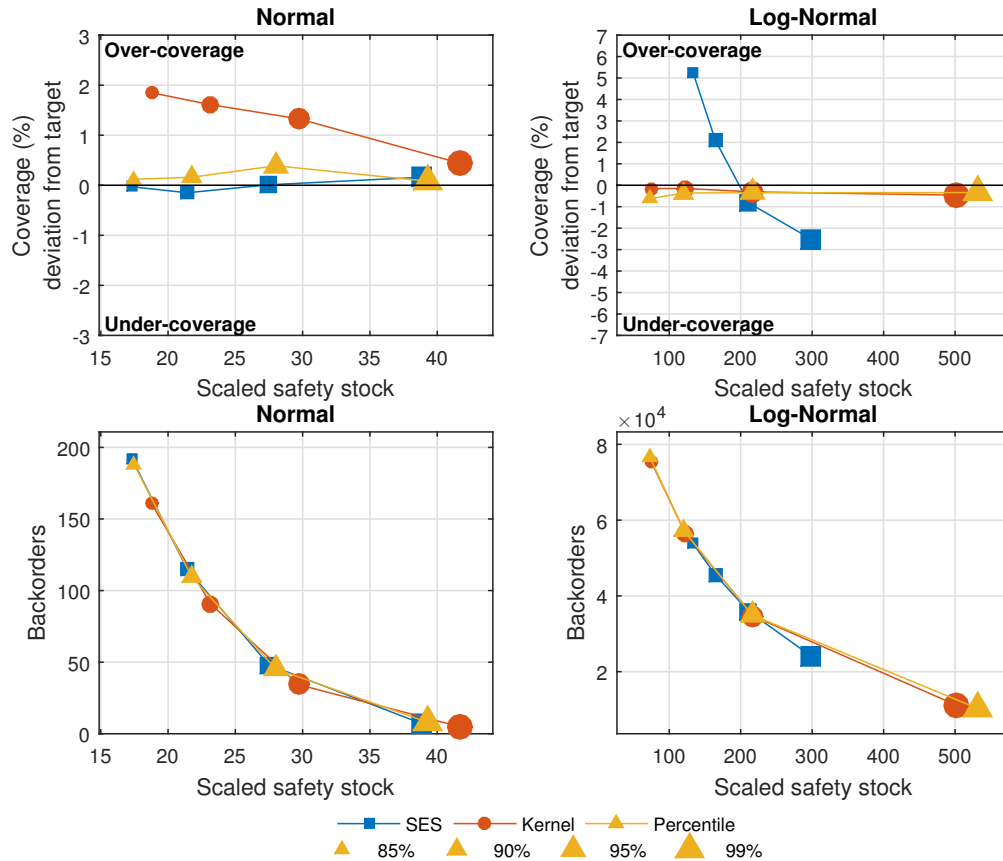


Figure 3: Trade-off curves for a sample size of 500 observations. Upper and lower left panels: Normal distribution. Upper and lower right panels: Log-normal distribution.

kernel provides again a higher coverage rate than the targeted.

Summarizing, non-parametric methods like Kernel and percentile can provide better trade-off curves for log-normal distributions. However, when the sample size is small their performance deteriorates. If the distribution is normal, parametric approaches work well as it is expected, whereas Kernel provides over-coverage and Percentile is more sensitive to small sample sizes. Given that these simulations do not consider correlations of the forecast errors or changes of volatility, the GARCH model do not show any advantage and may even present estimation problems.

5.2. Demand with time-varying volatility

In this simulation we have carried out two experiments with time-varying volatility. In the first case, the demand follows a normal distribution with constant mean ($\mu = 150$), and two different standard deviations ($\sigma_1 = 25$ and $\sigma_2 = 50$) with a sample size of 500 observations. σ_1 has been employed for time periods corresponding to 1:100, 201:300 and 401:500. This is done to have volatility changes in both the in-sample and hold-out sample. In the second case, a demand with constant mean set to 50 plus a stochastic term that follows a GARCH(1,1) model with parameters ($\omega = 0.01$, $a_1 = 0.4$, $\beta_1 = 0.5$) has been simulated. Both experiments have been repeated 100 times.

Figure 4 shows the 100 realizations of the simulated demand, where the effect of the time-varying standard deviation is apparent. Upper plot in that figure shows the simulated demands with volatility abrupt changes. The lower plot depicts the GARCH(1,1) realisations.

Figure 5 depicts the average trade-off curves obtained by each technique considered for both experiments. The right panels of that figure shows the trade-off curves for the first experiment with volatility abrupt changes. Regarding backorders, SES and GARCH yield slightly lower backorders for similar scaled safety stocks. With respect to coverage, SES and GARCH provides over-coverage for lower targets and under coverage for the highest target. Kernel produces over-coverage for each target, which it is reduced for larger target coverages. Interestingly, Percentile provides a coverage deviation close to zero.

The left panels of that figure shows the trade-off curves for the second experiment based on a GARCH simulation. In general terms, the GARCH provides low coverage deviations at a low scaled safety stock. Additionally, the level of backorders provided by parametric approaches is lower than the non-parametric counterparts even for lower scaled safety stocks. This result can be important, since the level of backorders can be used as a signal of potential autocorrelation in the forecast errors.

5.3. Influence of the lead time

The lead time in previous simulations was set to 1. Now we study the influence of the lead time on the safety stock computation. The lead time was set to 4 periods. We also set the sample size to 500 observations and use a demand that follows an ARIMA(0,1,1) with $\theta = -0.75$ and standard deviation $\sigma = 2$. We have used that ARIMA model because the single exponential

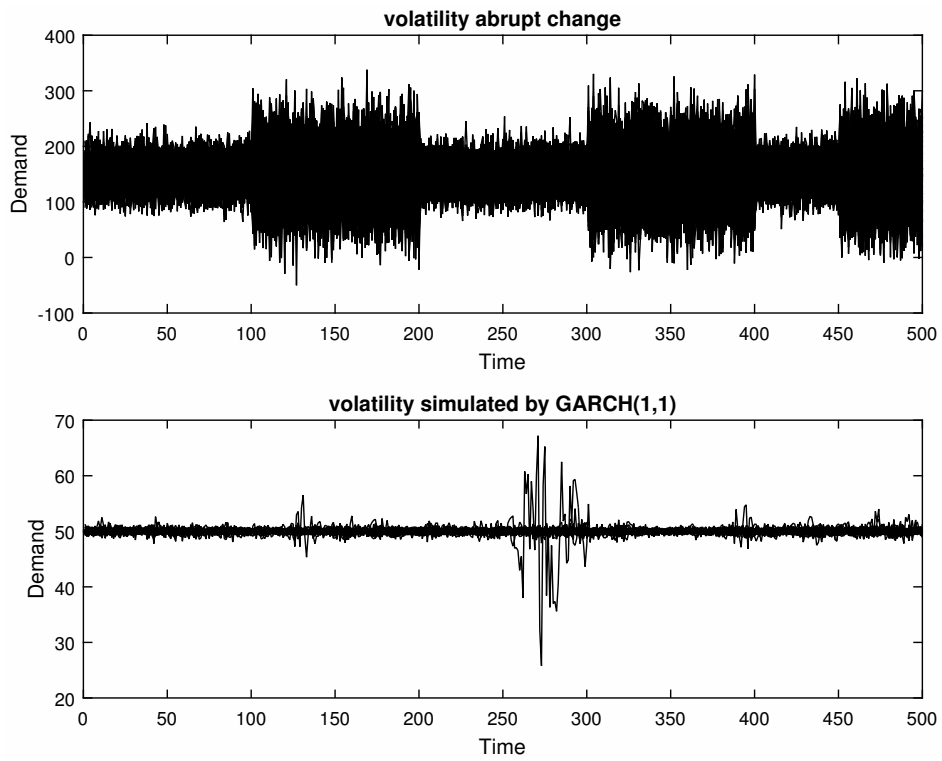


Figure 4: Demand realisations of the 100 experiments with time-varying volatility. Upper plot: Experiment with volatility abrupt change. Lower plot: GARCH(1,1) simulation

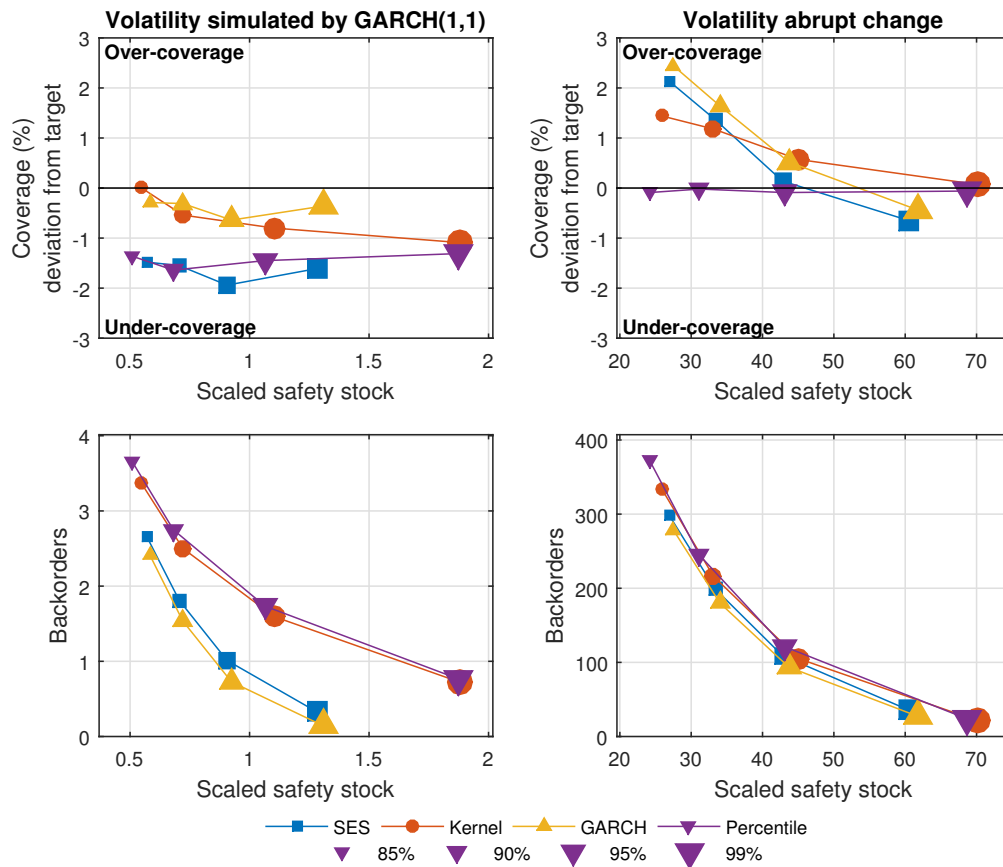


Figure 5: Trade-off curves for the experiments shown in Figure 4

smoothing is optimal for such a model and thus, all the differences between the methods considered are independent of the point forecasting model. In fact, the value of $\theta = -0.75$ corresponds with a theoretical $\alpha = (1+\theta) = 0.25$. To have positive demand values a constant of 50 units was added to the signal. Given that the lead time is greater than 1, we can add to the previous empirical approaches the theoretical approaches from equations (8) and (9), which are denoted by $\sigma_L(8)$ and $\sigma_L(9)$, hereafter. Note that methods like SES, Kernel, GARCH and Percentile are based on the empirical lead time standard deviation from (10)

Figure 6 shows that $\sigma_L(8)$ provides lower coverage and a higher number of backorders than $\sigma_L(9)$ given the simplifications assumed by $\sigma_L(8)$. Percentile, Kernel, GARCH and $\sigma_L(9)$ achieve coverages very close to the target ones, although Kernel provides coverages higher than the targeted at a higher scaled safety stock. Interestingly, regarding the coverage, SES performs worse than the non-parametric approaches, although GARCH performs well. Regarding backorders, GARCH obtains a very good performance even better than $\sigma_L(9)$. Recalling the time-varying simulations in the previous section, we concluded that when a parametric model like GARCH provided lower backorders than the rest of the methods it could indicate possible autocorrelation. Figure 6 shows a better performance from GARCH, what possibly indicates that there is some autocorrelation that has not been considered in the theoretical expression (9) and that GARCH can incorporate empirically. We conjecture that such a potential autocorrelation may be induced by the overlapping temporal demand aggregation, since our results agree with the conclusions obtained by (Boylan and Babai, 2016), where the overlapping approach reduces the inventory backordering volumes, whilst maintaining the same inventory holding volumes.

5.4. Stock control policies

In the previous section, we focused on measuring the performance of prediction intervals, avoiding the use of a specific stock control policy. So far their good or poor performance did not depend on the stock control policy. Nonetheless, it is important to show how these techniques can be implemented along with a stock control policy and to relate to the results above, for the prediction interval metrics (coverage, scaled safety stock and backorders), with conventional stock policy metrics (service level, inventory investment and backorders). We incorporate an order up to stock control policy following the work by (Dejonckheere et al., 2003), shown in Figure

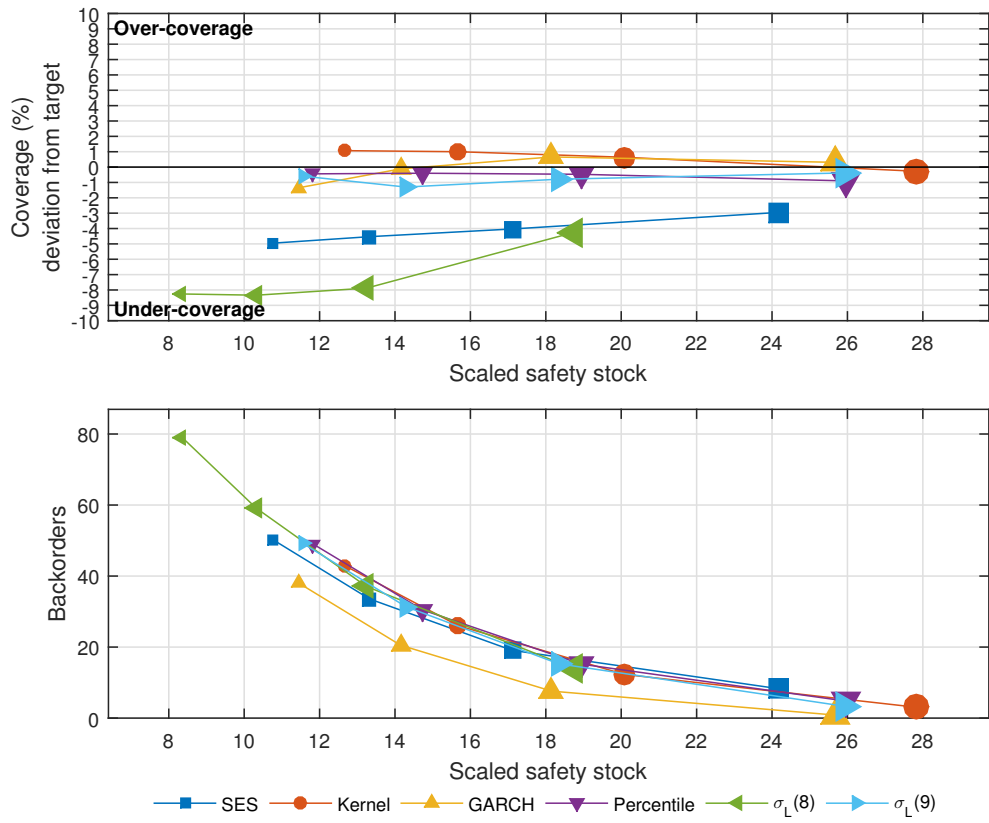


Figure 6: Trade-off curves for a demand that follows an ARIMA(0,1,1) and a lead time of 4 periods

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

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

$$O_t = S_t - IP_t \quad (19)$$

where O_t , S_t and IP_t stand for the ordering decision, order-up-to level and inventory position at time t , respectively. The inventory position can be defined in terms of the net stock (NS) and outstanding orders (OO) as:

$$IP_t = NS_t + OO_t \quad (20)$$

where NS_t equals inventory on hand minus backorders. We use the variable NS_t in order to compute the typical stock control metrics. In other words, the customer service level is calculated as the proportion of times that the NS is greater than zero. Backorders are calculated as the sum of negative values of NS_t across time and subsequently averaged across SKUs, and finally, inventory investment is the NS_t average across time and SKUs.

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

$$S_t = F_L + k\sigma_L \quad (21)$$

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 coincides with the upper bound of the prediction interval defined in (18).

The main difference of the block diagram model shown in Dejonckheere et al. (2003) and the one shown in Figure 7 is for σ_L . In Dejonckheere et al. (2003) the σ_L is not computed and the uncertainty is considered by inflating the lead time in one unit. In our work, we focus on the different ways to compute σ_L and how it affects at the safety stock performance. For this simulation we use expression (9) to compute σ_L , although any of the methods presented can be used, since we are analyzing the links between prediction interval and stock control metrics. The described simulation is implemented in SIMULINK.

Figure 8 shows the relationship between the prediction intervals performance metrics and the conventional stock control metrics. In the three plots, the x-axis represent the stock control metrics and the y-axis the prediction

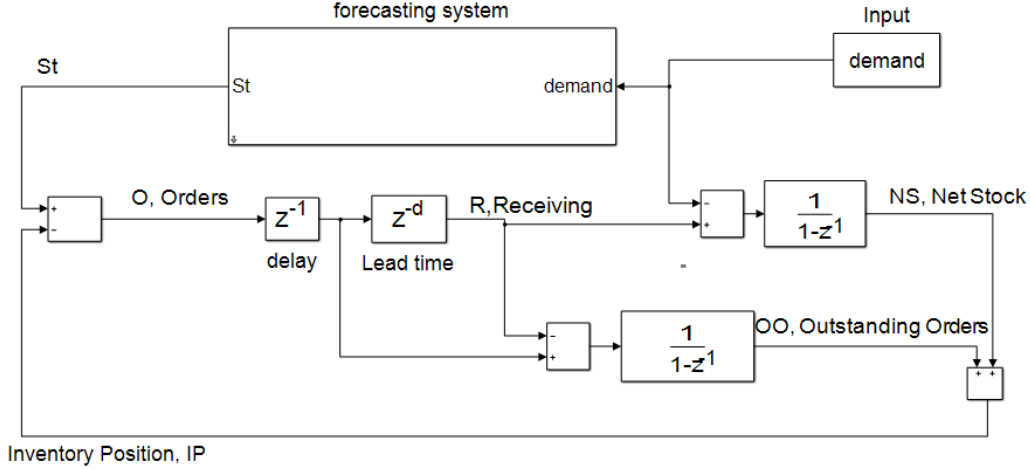


Figure 7: Order-up-to stock control policy plus forecasting system based on (9)

interval metrics. We can appreciate a linear relationship in all of them. In other words, the use of prediction intervals metrics is a good approximation to predict what is going to be the stock control metrics, at least for an order-up-to stock control policy.

6. Case study data

The data employed in this paper has been previously used by (Barrow and Kourentzes, 2016). The data comes from a major UK fast moving consumer goods manufacturer specialized in the production of household and personal care products. In total 229 products with 173 weekly observations per product are available. According to (Barrow and Kourentzes, 2016) the SKUs did not present seasonality and about 21% of them presented a trend.

The data is split into three parts. The first part (20% of the data) is used to compute both the α_1 exponential smoothing parameter and the initial value to determine the point demand forecast. The second part (60% of the data) is employed to optimize the volatility models (parametric and non-parametric). That second part is larger than the first one to prevent non-parametric percentile model sample size issues and to obtain a fair comparison. Finally, the last part (20% of the data) is devoted to test the probabilistic forecasts of the alternative methods. We employ single exponential

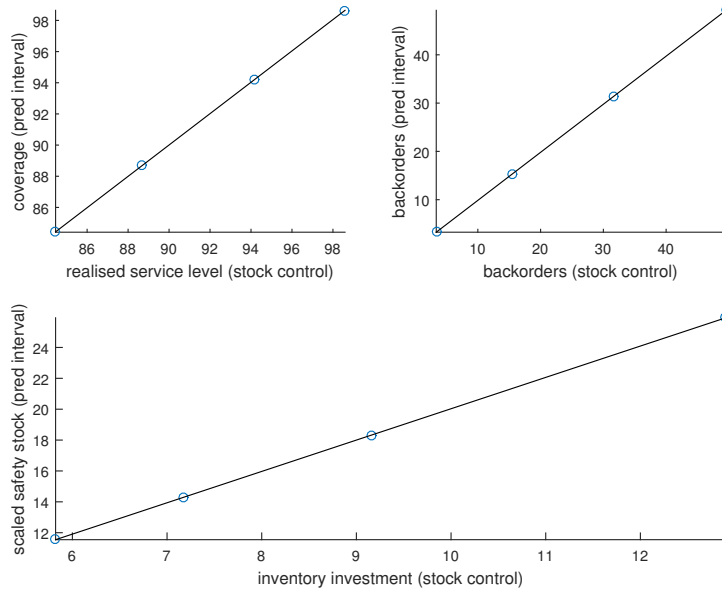


Figure 8: Relationship between prediction interval metrics and stock control policy metrics

smoothing to produce the point forecasts, although some SKUs were reported to potentially include a trend. We use the Augmented Dickey-Fuller test to find that first differences are adequate for these series and therefore single exponential smoothing is still appropriate. Note that trend exponential smoothing (Holt’s linear trend method) is equivalent to ARIMA(0,2,2), and therefore is not needed in this case.

Figure 9 shows the trade-off curves of the manufacturer data for a lead time of 4 weeks. Considering the theoretical approaches ($\sigma_L(8)$ and $\sigma_L(9)$), $\sigma_L(9)$ obtains a lower coverage deviation and a lower level of backorders than $\sigma_L(8)$, as it was expected. Regarding the empirical non-parametric approaches (Kernel and Percentile), Kernel provides a lower coverage deviation from the target and also a lower level of backorders with a higher scaled safety stock. Finally, the empirical parametric approaches (SES and GARCH) show a very good performance. This is particularly true for the GARCH model. It is clear that the lowest coverage deviation among all the methods is obtained by GARCH, at a similar level of scaled safety stock. Additionally, it performs very well with respect to backorders. Note that this better performance of GARCH indicates that SKUs from this dataset, in general terms, present important autocorrelations on the lead-time forecast errors. Thus, the use of

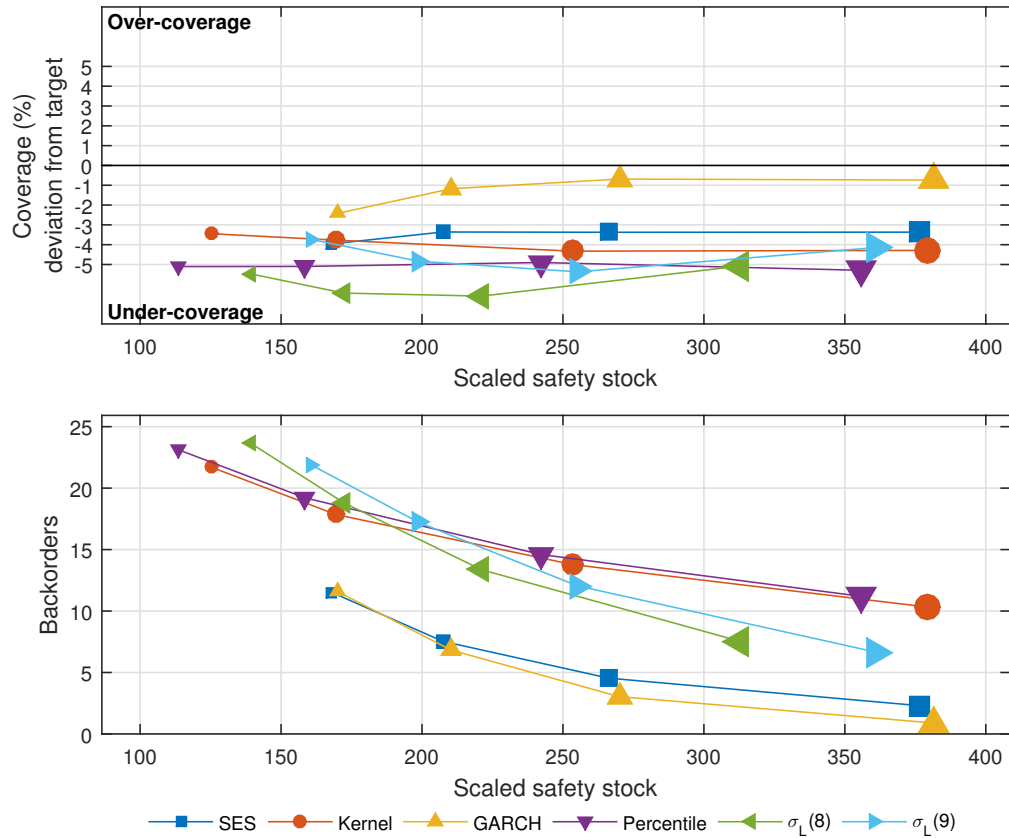


Figure 9: Trade-off curves for the manufacturer data assuming a lead time equal to 4 weeks

GARCH may bring significant economic savings.

7. Conclusions

Despite the attention from both academics and practitioners on supply chain risk management, the links between demand uncertainty and risks are still under-research. One tool that supply chains typically employ to prevent further risks is the safety stock. This work examines empirical approaches, parametric and non-parametric, to estimate the variability of the forecast errors in order to determine the appropriate safety stocks. In addition, those empirical methods are compared to traditional theoretical methods described in the supply chain literature. Our intention has been to provide recommen-

dations for the case that the assumption of normal i.i.d. forecast errors does not hold, which is the norm in practice.

Simulation results show that empirical non-parametric approaches like Kernel can be well-suited when the forecast error statistical distribution is fat tailed, i.e., the normality assumption is violated. Additionally, if the independence assumption cannot be guaranteed, empirical parametric approaches like SES and GARCH offer a promising alternative. In fact, when forecast error are autocorrelated, GARCH and SES perform well. Moreover, we find that when GARCH or SES improve the level of backorders it might indicate potential autocorrelation. Under such circumstances GARCH model is capable of capturing this and provide a better performance with respect to forecast coverage and particularly for backorders volume.

Experimental results using real data are also provided. These validate the previous conclusions obtained by the simulations exercise. GARCH models, in comparison with the rest of techniques, provided the best forecast coverage with a similar safety stock and the lower level of backorders. Recall that this latter feature could indicate that this data set has autocorrelation on the lead time forecast errors.

These results have been assessed by means of prediction interval metrics, like forecast coverage and prediction interval width, which have been redefined as scaled safety stock. Their implications on traditional stock control metrics like customer service level, inventory investment and backorders have been analyzed by including a simulation with an order-up-to level stock policy. The results showed a high correlation between prediction interval metrics independent of the stock control policy and common stock control metrics.

Further research should address some limitations of this work. Basically, we have analyzed single exponential smoothing and GARCH(1,1) model for point and volatility forecasts, respectively. Future works should incorporate automatic identification routines to better identify optimal point and volatility forecasting models, apart from the ones studied here. Moreover, 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 the project with reference DPI2015-64133-R.

References

- Axsäter, S., 2007. Inventory Control. International Series in Operations Research & Management Science. Springer US.
- Barrow, D., Kourentzes, N., 2016. Distributions of forecasting errors of forecast combinations: implications for inventory management. *International Journal of Production Economics* 177, 24–33.
- Beutel, A.-L., Minner, S., 2012. Safety stock planning under causal demand forecasting. *International Journal of Production Economics* 140 (2), 637 – 645.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31 (3), 307 – 327.
- Boudoukh, J., Richardson, M., Whitelaw, R. F., 1997. Investigation of a class of volatility estimators. *The Journal of Derivatives* 4 (3), 63–71.
- Boylan, J. E., Babai, M. Z., 2016. On the performance of overlapping and non-overlapping temporal demand aggregation approaches. *International Journal of Production Economics*, –In press.
- Chatfield, C., 2000. Time-series forecasting. CRC Press.
- Dejonckheere, J., Disney, S. M., Lambrecht, M. R., Towill, D. R., 2003. Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research* 147, 567–590.
- Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society* 50, 987–1007.
- Eppen, G. D., Martin, R. K., 1988. Determining safety stock in the presence of stochastic lead time and demand. *Management Science* 34 (11), 1380–1390.

- Fahimnia, B., Tang, C. S., Davarzani, H., Sarkis, J., 2015. Quantitative models for managing supply chain risks: A review. *European Journal of Operational Research* 247 (1), 1 – 15.
- Gardner, E. S., 1985. Exponential smoothing: The state of the art. *Journal of Forecasting* 4 (1), 1–28.
- Gardner, E. S., 2006. Exponential smoothing: The state of the art, Part II. *International Journal of Forecasting* 22, 637–666.
- Gardner, Jr., E. S., 1990. Evaluating forecast performance in an inventory control system. *Management Science* 36 (4), 490 – 499.
- Graves, S. C., 1999a. Addendum to "a single-item inventory model for a non-stationary demand process". *Manufacturing & Service Operations Management* 1 (2), 174–174.
- Graves, S. C., 1999b. A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* 1 (1), 50–61.
- Heckmann, I., Comes, T., Nickel, S., 2015. A critical review on supply chain risk - definition, measure and modeling. *Omega* 52, 119 – 132.
- Heizer, J., Render, B., 2004. *Operations Management*. No. v. 1 in *Operations Management*. Pearson Prentice Hall.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag, Berlin.
- Isengildina-Massa, O., Irwin, S., Good, D. L., Massa, L., 2011. Empirical confidence intervals for usda commodity price forecasts. *Applied Economics* 43 (26), 3789–3803.
- Katsikopoulos, K., Syntetos, A., 2016. Bias-variance trade-offs in demand forecasting. *Foresight: The International Journal of Applied Forecasting* (40), 12–19.
- Kourentzes, N., 2014. On intermittent demand model optimisation and selection. *International Journal of Production Economics* 156, 180 – 190.

- Lee, Y. S., Scholtes, S., 2014. Empirical prediction intervals revisited. *International Journal of Forecasting* 30 (2), 217 – 234.
- Nahmias, S., 2009. *Production and Operations Analysis*, sixth Edition. Operations and decision sciences. McGraw-Hill/Irwin.
- Nelson, D. B., 1990. Stationarity and persistence in the garch(1,1) model. *Econometric Theory* 6 (3), 318–334.
- RiskMetrics, 1996. Riskmetrics technical document. Tech. rep., J.P. Morgan/Reuters.
- Silver, E., Pyke, D., Peterson, R., 1998. *Inventory Management and Production Planning and Scheduling*. Wiley.
- Silverman, B., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, Bristol.
- Syntetos, A. A., Babai, M. Z., Jr., E. S. G., 2015. Forecasting intermittent inventory demands: simple parametric methods vs. bootstrapping. *Journal of Business Research* 68 (8), 1746 – 1752, special Issue on Simple Versus Complex Forecasting.
- Tashman, L., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16, 437–450.
- Taylor, J. W., 2004. Volatility forecasting with smooth transition exponential smoothing. *International Journal of Forecasting* 20 (2), 273 – 286, forecasting Economic and Financial Time Series Using Nonlinear Methods.
- Trapero, J. R., Kourentzes, N., Fildes, R., Feb 2015. Identification of sales forecasting models. *J Oper Res Soc* 66 (2), 299–307.
- Wecker, W. E., January 1979. The variance of cumulative demand forecasts. Working Paper. Graduate School of Business. University of Chicaco 5.
- Williams, W. H., Goodman, M. L., 1971. A simple method for the construction of empirical confidence limits for economic forecasts. *Journal of the American Statistical Association* 66 (336), 752–754.