Computers & Industrial Engineering 142 (2020) 106380

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion

Mahdi Abolghasemi[a], Eric Beh[a], Garth Tarr[b], Richard Gerlach[c]

[a] School of Mathematical and Physical Sciences, The University of Newcastle, NSW, Australia
[b] School of Mathematics and Statistics, The University of Sydney, Sydney, NSW, Australia
[c] The University of Sydney Business School, Sydney, NSW, Australia

ARTICLE INFO

Keywords:
Demand volatility
Promotions
Forecasting models
Robust forecasts
Inventory performance

ABSTRACT

The demand for a particular product or service is typically associated with different uncertainties that can make them volatile and challenging to predict. Demand unpredictability is one of the managers' concerns in the supply chain that can cause large forecasting errors, issues in the upstream supply chain and impose unnecessary costs. We investigate 843 real demand time series with different values of coefficient of variations (CoV) where promotion causes volatility over the entire demand series. In such a case, forecasting demand for different CoV require different models to capture the underlying behavior of demand series and pose significant challenges due to very different and diverse demand behavior. We decompose demand into baseline and promotional demand and propose a hybrid model to forecast demand. Our results indicate that our proposed hybrid model generates robust and accurate forecast and robust inventory performance across series with different levels of volatilities. We stress the necessity of decomposition for volatile demand series. We also model demand series with a number of well known statistical and machine learning (ML) models to investigate their forecast accuracy and inventory performance empirically. We found that ARIMA with covariate (ARIMAX) works well to forecast volatile demand series, but exponential smoothing with covariate (ETSX) has a poor performance. Support vector regression (SVR) and dynamic linear regression (DLR) models generate robust forecasts across different categories of demands with different CoV values. In terms of inventory performance, ARIMAX and combination models have superior performance to the other presented models. The hybrid algorithm also depicts robust performance across different series with different CoVs and has low inventory costs.

1. Introduction

Demand forecasting is the basis for many managerial decisions in the supply chain such as demand planning and order fulfillment Narayanan, Sahin, and Robinson (2019), production planning Donohue (2000), and inventory control Silver, Pyke, and Peterson (1998). It is usually difficult to perform demand forecasting at a high level of precision due to underlying volatility and many uncertainties Syntetos, Babai, Boylan, Kolassa, and Nikolopoulos (2016). Demand is one piece of important information that can be shared and used in supply chain management. Accurate and reliable demand forecasts provide vital intelligence for supply chain managers to support their planning and decision making.

Demand volatility inherently exists due to many endogenous and exogenous factors. Various variables such as promotion, weather, market trends, and season may have an impact on consumer behavior and contribute to demand volatility Gilliland (2010). Promotion, in particular, is a very common practice in the retailing industry that can induce demand volatility. The impact of promotion on demand dynamics has been investigated extensively in the literature (see, for example, Divakar, Ratchford, & Shankar (2005, 2014, 2017, 2016, 2015)). Demand volatility and variability can occur before, after or during promotional periods Blattberg and Neslin (1993).

Demand volatility is a challenging risk to the supply chain. Researchers and practitioners have repeatedly raised their concerns about increasing demand volatility as a threatening risk to the supply chain Christopher and Holweg (2017). Demand volatility renders demand forecasting a difficult task and poses excess costs for stock-outs, inventory, and capacity utilisation Christopher and Holweg (2011). However, demand volatility has been under-considered in supply chain demand forecasting and inventory forecasting literature. Forecasting to capture the underlying behavior of the volatile demand is crucial to diminish the uncertainty in different levels of the supply chain and reduce the costs.

We use the coefficients of variation (CoV) to measure volatility and propose appropriate demand forecasting models when demand series exhibit different values of CoV. The CoV is the sample standard deviation divided by the sample mean. By definition CoV is a scale-independent metric and can be used to compare the relative variation across multiple series. It represents the uncertainties in data and it is

https://doi.org/10.1016/j.cie.2020.106380
Received 31 July 2019; Received in revised form 6 February 2020; Accepted 20 February 2020
Available online 22 February 2020
0360-8352/

considered as the demand variability criteria in supply chain Cachon (1999). We use this metric to measure the promotional variations in demand as well as the natural variations in demand. We consider the entire series since promotion not only impacts demand over the promotion periods, but it can affect demand in the periods before and after promotions. In general, demand series with large CoV values are associated with more uncertainties and are difficult to forecast Huang, Chang, and Chou (2008). These demands, if not forecasted accurately, can cause many problems in the upstream supply chain operations such as inventory control and pose a shock to supply chain performance. One aim of this paper is to find the most suitable forecasting model for demand series with different levels of volatility.

Time series for demand with large CoV are associated with different uncertainties and are difficult to forecast (Huang et al., 2008). While there is no single model that performs well for all different types of demand series, we can identify the appropriate forecasting model based on time series features Wang, Smith, and Hyndman (2006, 2009). Recommendations regarding suitable forecasting methods for time series that exhibit various features have been recommended by researchers in different contexts Armstrong (2001, 2014, 2009). More specifically in supply chain context, Shah (1997), Meade (2000) proposed CoV as one of the features in selecting the appropriate forecasting model. Yet there has been a lack of theoretical and empirical studies undertaken that investigate the volatility of demands that are impacted by promotion as a criterion to develop a forecasting model and improve supply chain performance.

The main contributions of this study are twofold: methodological and practical. Firstly, the methodological contribution is to introduce CoV as a criterion that represents the volatility of demand in a supply chain context. We argue that the CoV for a demand series in a supply chain context is representative of the variability of a time series and of features such as entropy, kurtosis, and skewness. Secondly, we develop and implement a number of well-known forecasting methods and evaluate their accuracy when demand series exhibit different degrees of volatility and show that demand volatility impacts the performance of forecasting models. We argue that demand series with different CoVs require different forecasting methods that suits their characteristics. Thus, the CoV should be considered in developing appropriate forecasting models and, consequently, generating robust forecasts. Thirdly, we show that a decomposition-based model works well to forecast the volatile demand. We decompose demand to the non-promotional (baseline) and promotional demand and propose a hybrid model that uses a piecewise regression model and a time series model to forecast the promotional and baseline demand, respectively. We compare this model with other well-known models in the literature and show that our model has several advantages to the other existing models in the literature in terms of accuracy, simplicity, robustness to volatility, and inventory performance.

We also make a number of practical contributions. First, considering the CoV of demand series, managers will be able to use an appropriate model to capture the volatility of demand with a high level of accuracy. Second, we develop a forecasting model that only uses price as the explanatory variable. This model is simple, intuitive, and easy to use in practice, so makes our procedure desirable for practitioners Armstrong (2001). Third, we evaluate the inventory performance of models and show that demand volatility can significantly impact the inventory costs and performance. Managers can choose their desirable model to reduce their inventory level, use their sources more efficiently and potentially improves the performance of the entire supply chain.

The remainder of the paper is organised as follows. Section 2 reviews the literature. Section 3 discusses the methodology and forecasting techniques that are used to forecast demand. Section 4 describes our data set and Section 5 provides the empirical results and discussion. Conclusions are drawn in Section 6.

2. Literature review

Supply chains are associated with many uncertainties and complexities in the modern world. These uncertainties arise for many reasons such as the activities of partners, customers behavior, competitors behavior, emerging technologies, and new product development; all of which contribute to a volatile supply chain and a volatile demand Christopher and Holweg (2011, 2013). In order to avoid the negative consequences of volatile demand, it is important to incorporate and predict these uncertainties.

The accuracy of demand forecast is one of the factors that can contribute to the volatility of demand Gilliland (2010). Demand forecasting is not an easy task and many companies and forecasters fail to undertake a scientific forecast Armstrong and Green (2017). The biggest problem with demand forecasting is the uncertainty in demand that renders demand forecasting a challenging problem Syntetos et al. (2016). Accurate demand forecasting is a critical factor in determining the quality of decision making. Inaccurate forecasts may cause unnecessary costs in procurement and transportation, manpower, service level, and inventory Torkul, Yılmaz, Selvi, and Cesur (2016). While demand volatility can be reduced and controlled, it is inevitable. Thus, it is crucial to have a proper strategy to control volatility.

Researchers and practitioners have proposed and developed techniques and approaches to deal with demand volatility. For example, one way to avoid the negative impact of demand volatility is to increase the inventory level. This will help to counter the demand volatility but imposes a lot of costs to the companies Meindl and Chopra (2001). Another strategy to control demand volatility is to increase the capacity, but this is not an attractive proposition as it costs a lot for the supply chain. While these approaches can help address issues associated with demand volatility, they may not be cost-effective.

Demand forecasting is a prerequisite for strategies that aim to control the volatility of demand Hope and Fraser (2003). It is the first step towards dealing with uncertainty and volatility in the supply chain. Over the past few decades, many different models have been used for retail sales forecasting Aye, Balcilar, Gupta, and Majumdar (2015, 2008, 2014, 2014, 2017, 2016, 2005, 2016). There is no unique solution that can address all types of forecasting problems and performs better than all other forecasting models. Nor are there any specific methods that perform better than all other methods in all possible situations and under all possible conditions. However, some models might outperform others under particular conditions. For instance, Alon, Qi, and Sadowski (2001) compared the forecasts generated by an artificial neural network (ANN), triple exponential smoothing, ARIMA, and multiple aggregation methods. They concluded that ARIMA and triple exponential smoothing outperform the other models when the macroeconomic conditions are stable, whereas ANN and multiple aggregations might work better in volatile markets. Huang et al. (2014) considered the values of competitive information such as price and promotion in demand forecasting. They first identified the most relevant explanatory variables and then developed an autoregressive distributed lag model that generates promising results. In another study, Huang, Fildes, and Soopramanien (2019) proposed methods that take into account the structural change in demand caused by marketing activities. These methods underestimate or do not explicitly investigate volatility as a factor that can impact the performance of the forecasting models.

Extrapolative methods are popular forecasting methods that assume the past pattern of a series is representative of its future. Therefore, we can estimate the future of the series based on its relationship with other variables. They have been applied to various forecasting problems and can be used as benchmark models Makridakis, Spiliotis, and Assimakopoulos (2018). We use ARIMA and exponential smoothing as two of the popular and powerful extrapolative models Hyndman (2018, 2014). These methods rely on auto-correlations with past data and base their forecasts on extrapolations from past patterns. Hence, they work well only when the future behaves similarly to what have been observed in the past; however, pattern changes can severely compromise their efficacy. Conventional time series models lack the ability to adequately capture random variation in series and ignore the impact of influencing variables on demand Huang et al. (2008). Therefore, a univariate ARIMA or exponential smoothing model might fail to forecast well if demand time series is subject to volatility. To overcome this problem, one can add influential variables to a time series and construct ARIMAX and ETSX models. These models consider both autocorrelations and influential factors and are promising approaches to the time series literature. We also implement these models as alternative benchmark methods that take into account the promotion impact.

Econometric forecasting models have been developed to address the pitfalls of conventional extrapolative forecasting models. Causal and regression models are a very popular form of econometric models. Various forms of causal models have been developed for retail sales forecasting Ali, Sayın, Van Woensel, and Fransoo (2009). More recently, other sophisticated regression models such as SCAN*PRO Van Heerde, Leeflang, and Wittink (2002) and CHAN4CAST Divakar et al. (2005) have been proposed to capture promotion impact. These models are static and designed to work at the brand level. The downside of these models is that they are expensive to use as they need many inputs variables. Moreover, it is not clear if these models perform robustly to forecast volatile demand across the entire demand series and not only during the promotion as a special event. The retail market is a dynamic environment in which the demand time series is changing continuously due to many different factors of major and minor influence. Therefore, a dynamic approach may fit and forecast the volatile demand better.

Demand series with high volatility may take different values over the horizon. One can use different models to forecast similar parts of demand series more accurately, and then combine them to forecast the entire demand series. Hybrid models that use forecast combination approaches have been successfully implemented in retail sales forecasting Aye et al. (2015, 2017, 2005). We can use different methods for combining forecasts Clemen (1989). The combination can be used at the beginning of modeling by decomposing a series into its components and constructing a hybrid model. Alternatively, the forecaster may use individual models to forecast a series and take an average from the generated results. An empirical study of the retail industry has shown that a hybrid ARIMA and ANN model improves the accuracy of demand forecasts Aburto and Weber (2007). In other research, a hybrid SVR and ARIMA model outperformed the standard statistical models Pai and Lin (2005). Theta is one of the hybrid models based on the decomposition of a time series and has attracted a great deal of attention after it was successfully implemented on M3-competition data Makridakis and Hibon (2000). This model decomposes the data into two Theta lines. The first Theta line removes the curvature of data and the second line doubles the curvature of data. The mean of these two models generates the final forecast Assimakopoulos and Nikolopoulos (2000). We develop a hybrid model that first decomposes demand into two components (baseline demand and uplift demand), forecasts them separately, and then combines them to predict the total demand.

During the last couple of decades, machine learning (ML) algorithms have gained a lot of attention in demand forecasting Abolghasemi, Hyndman, Tarr, and Bergmeir (2019b, 2018). However, they have not been fully explored in a supply chain management context and require further attention Min (2010). ML algorithms are computationally expensive but provide a range of different and flexible models for forecasting demand. Much debate surrounds the relative performance of statistical and ML methods, and it is difficult to draw general conclusions about their efficacy Aye et al. (2015, 2018). Each class of models might outperform others under certain conditions Ahmed, Atiya, Gayar, and El-Shishiny (2010). We utilise ANN and SVR as two of the most common and successful ML techniques for retail sales forecasting Makridakis et al. (2018, 2010). The performance of these models has not been fully explored when they face volatile demand time series. We apply these models to our data set and compare the results obtained using these techniques with those obtained using a number of statistical models.

We also empirically investigate a large data-set of demand time series that have demands with different levels of volatilities to develop a forecasting model. We aim to improve the forecasting practice for volatile demand series while also giving managers insight into their inventory performance and their ability to cope with demand volatility in the supply chain. Furthermore, the empirical results obtained in this paper can be used to propose theories about demand volatility and inventory volatility in the ongoing development of supply chain forecasting models.

3. Methodology

The main aim of this paper is to better understand how demand volatility changes the efficacy of forecasting models, and how it can be modeled in the supply chain. We underpin our arguments with evidence from hundreds of demand time series of a major fast-moving consumer goods (FMCG) food company. We strive to gather as much data as possible to make sure the results are valid, robust and generalisable to other similar firms.

Many different variables can affect the dynamics of demand and it is difficult to distinguish their relative impact on demand. Given full information, we could build comprehensive multivariate models to forecast volatile demand. However, it may not be practical to gather all data or it might be expensive to collect. Moreover, excessive information may not be useful for modeling and forecasting Huang et al. (2014). As such, it is of particular importance to choose the most influential variables that are available and are believed to drive the outcome of demand Ali et al. (2009, 2016). In the current study, we do not have access to all of the variables that may contribute to the volatility of demand. We only have access to the price of all products. The average of Pearson's correlation between price and demand is -0.83 indicating a very strong negative relationship that can potentially explain a high portion of demand variation. Thus, we rely on price as an influential factor to construct and evaluate different techniques for demand modeling. We propose a hybrid regression time series model, called HR-ARIMA. We also build different models from well-established methods in the literature including ARIMAX, ETSX, DLR, ANN, and SVR. We discuss these models in detail in the following subsections.

3.1. Hybrid model

The hybrid model was formulated to reflect the fact that not only the behavior of demand during promotional periods is completely different from that during non-promotional periods but also demand varies significantly during promotions and so, they require different models to accommodate this difference. Therefore, we decomposed demand into two main factors; these are referred to as baseline demand and promotional demand. One can decompose promotional demand more in detail, however, in this paper, we are interested only in the size of promotional demand and not the impact of different promotional factors on demand.

Understanding and estimating baseline demand is fundamental to many analyses in marketing and the basis for estimating the promotion behaviour. We used an ARIMA model to estimate baseline demand. In a similar manner to SCAN*PRO Van Heerde et al. (2002), we fitted an ARIMA model to the past non-promoted demand and used dummy variables to capture the promotional and post-promotional demands. The baseline demand is estimated in the absence of promotional variables. The models are implemented in R and the '*auto.arima*' function from the '*forecast*' package is used to fit and estimate the parameters of the ARIMA model Hyndman and Athanasopoulos (2014). Baseline demand is an estimate from the data and not an actual number. Therefore, it is difficult to measure whether the baseline demand estimate is accurate or not because there is no actual value to compare it with Gilliland (2010). Other simple, and more sophisticated methods, can be used to estimate the baseline demand Blattberg et al. (1995). We implemented ETS to forecast baseline demand as it is commonly used in industry and academia as a promising method (Wang & Petropoulos, 2016). The forecasting accuracy of ETS was found not to be significantly different to those described for the ARIMA model. Thus, we only report the ARIMA results.

After estimating baseline, demand uplifts are found simply by subtracting baselines demand from total demand at each period. The relationship between uplift demand and baseline demand can be either additive or multiplicative. We can transfer the multiplicative relationship into an additive relationship by calculating the natural logarithm of demand. Since we take the logarithm of demand, we estimate the size of demand uplift simply by subtracting the baseline demand from total demand. Surprisingly, we found that there is a stronger correlation between price and demand uplift only because of promotions than between total demand and price. Pearson's correlation between demand uplift only due to promotions and price is -0.89, indicating a strong linear relationship between price and demand uplifts.

Often a linear function is not appropriate to represent the relationship between sales and price for the entire sales data, but a ear function may not be suitable either. Thus, we use a linear regression model with the non-linear transformation of the variables to develop a piecewise regression model that can capture promotional demand. Piecewise regression is a type of polynomial regression where a number of regression models join together at knots. Knots are the points where the model parameter changes. In our case, knots are set on different existing promotional prices.

In Algorithm 1, a stepwise procedure is proposed for the construction of such a hybrid model:

Algorithm 1. Hybrid model algorithm

1: **for** $t = 0$ to N **do**
2: Decompose demand into main components that have lower volatility. This includes baseline demand and uplift in demand only due to the promotion.
3: Estimate the baseline demand for promotional periods.
4: Subtract baseline demand from total demand to find the uplift size due to promotion.
5: Set the promotional prices for each of the series to construct a piecewise regression model.
6: Construct a piecewise regression model for demand uplifts for different promotions.
7: **end for**
8: Forecast each of decomposed parts separately.
9: Sum both parts where appropriate.

We not only separated demand during promotions from demand during non-promotional periods but also decomposed the demand during promotional periods into baseline demand and demand occurring only due to the promotion. We then identified the different range of promotional prices for each product and fitted a piecewise regression model. The hybrid model can be shown using Eq. (1),

$$s_t = y_t + E(p) \tag{1}$$

where s_t denotes the total sales, $E(p)$ is the promotional sales, and y_t is the baseline demand at time period t.

This hybrid model has two components; an ARIMA, and a piecewise regression. The first component, y_t, is an modeled using an ARIMA process and could be replaced by any other model capable of forecasting baseline demands. The second component is a piecewise regression model which is used for forecasting demand uplifts during promotions Van Heerde et al. (2002). Note that, the piecewise regression model takes different coefficients for different promotional prices. It assumes that the promotion price is known in advance which is the case in many real world applications Armstrong (2001).

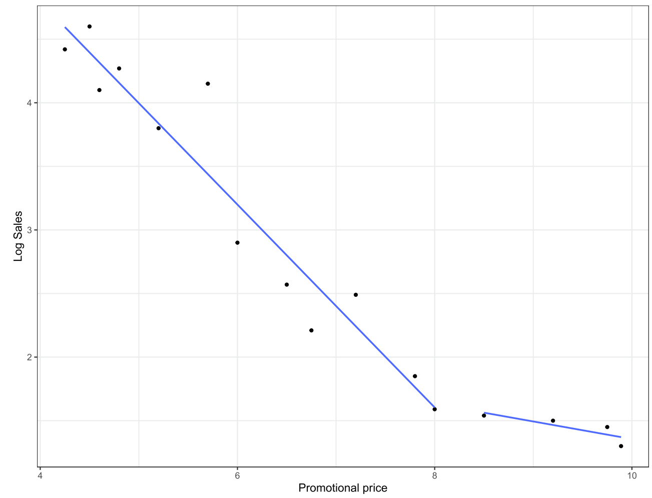

Fig. 1. Demands uplifts.

For example, Eq. (2) shows an expanded piecewise regression for the Fig. 1 when two ranges for promotions are offered,i.e., minor promotional prices when price is larger than $r_1 = 8$, and major promotions where price is smaller than $r_1 = 8$:

$$E(p) = \begin{cases} \alpha_1 + \beta_1 x & x \leqslant r_1, \\ \alpha_2 + \beta_2 x & x > r_1, \end{cases} \tag{2}$$

Note that, the knots are known in our case and set to the major ($x \leqslant r_1$) and minor ($x > r_1$) promotional prices for different SKULs. However, this may not be known in other cases. Then, it needs to be estimated. Interested readers can refer to Muggeo (2003) for details about estimating unknown knots.

3.2. ARIMAX

The ARIMAX model is a generalisation of the ARIMA model that can be constructed by adding an explanatory variable to the ARIMA model. The parameters p, d, and q in an ARIMA(p, d, q) model represent the autoregressive (AR) order component, the order of differencing and the order of the moving average (MA) component, respectively. Eq. (3) demonstrates an ARMA(p, q) model where e_t is a white noise process with mean zero and variance σ^2:

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}. \tag{3}$$

When additional covariate information is available, it is often beneficial to add it as a predictor to the forecasting model. ARIMAX can be constructed by adding price r_t as the covariate to the right hand side of Eq. (3), so that we have the model (4),

$$z_t = \beta r_t + \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q}, \tag{4}$$

where $z_t = \Delta^d y_t$.

The models are implemented in R using the '*arimax*' function in the '*TSA*' package to fit the ARIMAX model and estimate the parameters Chan and Ripley (2018). The models with the lowest AICc are chosen as the best fitted models.

3.3. ETSX

Exponential smoothing has frequently been applied in both research and practice since its introduction by Brown in 1959 Brown (1959). Exponential smoothing provides a weighted moving average in which the most recent observations are given more weight. Different types of exponential smoothing models employ different parameters, which enable them to capture irregular patterns. A statistical framework for exponential smoothing called ETS has recently been developed in the state space framework Hyndman, Koehler, Ord, and Snyder (2008). ETS methods are classified based on their components: errors, trends, and seasonality. The trend component includes a level (l) and a slope (b), which can be combined in different ways. In general, five types of trends, three types of seasonality, and two types of errors (additive and multiplicative) exist for ETS models, yielding a total of 30 different methods. The point forecasts for single and multiple sources of error are the same, but the prediction intervals are different Hyndman et al. (2008).

The following equations describe the ETS model,

$$y_t = l_{t-1} + \phi b_{t-1} + \varepsilon_t \tag{5}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t \tag{6}$$

$$b_t = \phi b_{t-1} + \beta\varepsilon_t. \tag{7}$$

Eq. (5) shows the observation, while Eqs. (6) and (7) show the state equations of the ETS model. Here, ε_t is a white noise process with mean zero and variance σ_ε^2.

When additional information is available, it can be embedded as a regressor into the ETS model. The ETSX model can be constructed by adding a covariate to the ETS to improve the forecasting accuracy. The regressor variable is added to the observation equation. In the following model, we have considered the price r_t as the covariate, and given it a time-invariant coefficient of c,

$$y_t = l_{t-1} + \phi b_{t-1} + cr_t + \varepsilon_t \tag{8}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t \tag{9}$$

$$b_t = \phi b_{t-1} + \beta\varepsilon_t. \tag{10}$$

Eq. (8) is the observation equation and Eqs. (9) and (10), with price at time t incorporated to the model, are the state equations of our ETSX model. Again, ε_t is a white noise process with mean zero and variance σ_ε^2.

The model and parameters are encoded and estimated in R using the '*es*' function in the *'smooth'* package Svetunkov (2018).

3.4. Dynamic linear regression

DLR is considered to be a generalisation of the standard linear regression model and can be expressed in a state space form. Although univariate models such as stochastic volatility models might provide quite good descriptions of the series behavior with irregularities or jumps; they will rarely be capable of predicting sudden changes without further information. DLR does not assume a regular pattern and stability of the underlying system but can accommodate for sudden and massive changes in demand Petris, Petrone, and Campagnoli (2009). Eq. (11) provides an observation equation for a DLR model which is initiated from the demand-price relationship. The state Eqs. (12) and (13) are the coefficients of a linear regression model which follows independent random walks Van Heerde et al. (2002),

$$\log(y_t) = \alpha_t + \beta_t \log(r_t) + \varepsilon_t \tag{11}$$

$$\alpha_t = \alpha_{t-1} + \nu_t \tag{12}$$

$$\beta_t = \beta_{t-1} + \omega_t, \tag{13}$$

where r_t denotes the price at time t, y_t represents the demand at time t, and ε_t, ω_t and ν_t are white noise processes with mean zero and variances σ_ϵ^2, σ_ω^2 and σ_γ^2, respectively.

Similar models with more variables can be used to forecast demand based on different variables. However, in our case, price plays a significant role and we can simplify the model to Eq. (11). Note that r_t is not stochastic. The DLR was implemented in R using the '*dynlm*' function in the *'dynlm'* package and parameters estimated with maximum likelihood Zeileis (2019).

3.5. Artificial neural network

An ANN is a supervised ML algorithm that is inspired by the human brain and learns from experience. An ANN consists of neurons and layers that are connected with arcs and input that are translated to output through an activation function. ANNs are powerful algorithms that are able to model any continuous, ear system, and can then make generalisations and predict unseen values Zhang, Patuwo, and Hu (1998). ANNs are used extensively in various fields of forecasting such as demand forecasting Aburto and Weber (2007).

In this paper, we employed the commonly used 'feed-forward error back-propagation' type of ANN with one hidden layer. In this type of ANN, the price variable is entered as the input to as inputs are entered to the neurons in each layer. Each neuron is connected to all of the neurons in the next layer. The aim is to minimise the error between the predicted and actual values of demand. It is vital to transfer the ANN data set before training the neural net. We transferred the data using the min–max method and considered price and demand time series as the inputs to forecast demand. We tested a sequence of different sizes by running the set size from 1 to 50 by increment size of 10. The learning rate decay was set based on a grid search on the sequence from 0.0 to 0.3 on a range of 0.1. We also tried both sigmoid and linear function and chose the sigmoid function as the activation function. The model was fitted using the *'nnetar'* function in the *'neuralnet'* package in R Günther and Fritsch (2010).

3.6. Support vector regression

SVR is a powerful supervised learning algorithm in which the output is a numerical variable and is the most common application of support vector machines (SVM) Basak, Pal, and Patranabis (2007). SVM implements intelligent algorithms to discover the patterns in complex data sets and has been used in many different applications including demand forecasting Villegas, Pedregal, and Trapero (2018).

The difference between SVR and ordinary least square regression models is that SVR attempts to minimise the generalised error, whereas statistical regression models try to minimise the deviation of forecasted values from the actual ones. SVR finds a linear function in the space within a distance of ϵ from its predicted values. Any violation of this distance is penalised by a constant, C Ali et al. (2009). SVR employs a kernel function to transfer low dimensional data to high dimensional data where the kernel function can take on a number of different forms. We tried different Kernel functions including poly, RBF, and a linear function. Finally, we used a grid search in order to choose the best combination of the kernel function, kernel coefficient (γ) and cost function C. We ran γ on a sequence of intervals of width 0.1 ranging from 0 to 1, and C on a sequence of intervals of width 1 ranging from 0 to 100. We fit the model using the *'svm'* function from the '*e1071*' package in R Dimitriadou et al. (2006) and selected the model that result in highest accuracy.

4. Data and time series features

We gathered data from a food manufacturing company producing hundreds of FMCG in Australia. Consumer demand data are available from the point of sales and prices are gathered by the company. Demand data are aggregated across the retailers and span 112 weeks. There are 843 stock keeping unit locations (SKULs). Demand levels differ greatly between promotional and non-promotional periods. This difference is mainly due to the promotion impact and promotion contributes significantly to the CoV.

Fig. 2 shows three different demand time series with different levels of CoV. These series have different features and depict different levels of volatility.

Demand during promotions can be up to 60 times greater than those during non-promotional periods for highly volatile demand series. The

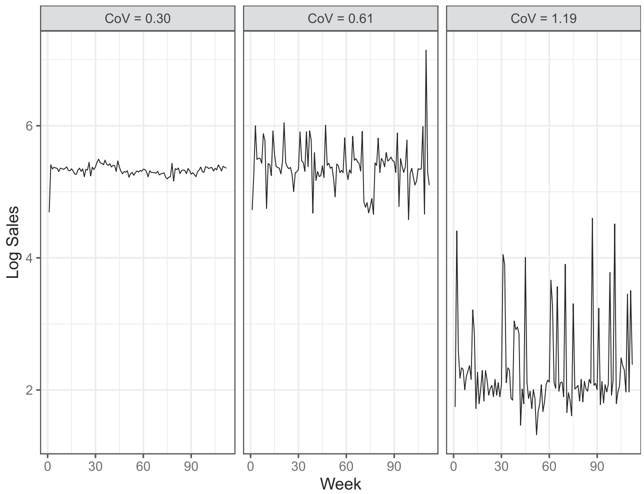

Fig. 2. Demands with different volatilities.

natural logarithm of demand has been employed to reduce the displayed variability of demand. Demands are heavily impacted by promotion but are not of a seasonally demanded nature. As mentioned previously, there is no unique best solution for forecasting all these series as a single model may not perform well for series with different features.

Since demand series are highly volatile, we have categorised them based on their CoV into three different groups. There are different categories to measure relative volatility of a time series Scholz-Reiter, Heger, Meinecke, and Bergmann (2012). We compute the CoV for training observations of each demand time series. There is no consensus on a cut-off value for CoV, rather it depends on the type of time series, related industry, and volume Syntetos, Boylan, and Croston (2005). Since our series are impacted by promotion, we opt for a larger threshold to classify series as volatile. We use the classification approach of Scholz-Reiter et al. (2012) to categorise the SKUL demands based on their CoV as follows:

- Group A: Low volatility demand where CoV is smaller than or equal to 0.5. There are 311 SKULs in this category. The average CoV of these demand series is 0.32.
- Group B: Moderate volatility where CoV is greater than 0.5 and smaller than or equal to one. There are 255 SKULs in this category where the average CoV of these demand series is 0.75.
- Group C: High volatility where CoV is greater than one. There are 277 SKULs in this category. The average CoV of these demand series is 1.71.

Demand time series has many different features that may define their behavior. Identifying and analysing these features may help us to find the most suitable forecasting model. We analyse the following important features for each group of series: stationarity, trend, seasonality and seasonal strength, skewness, kurtosis, entropy, and non-linearity. These features are defined and calculated as follows:

1. Stationarity: A time series is stationary if it has constant mean and variance across the horizon, and autocovariance does not depend on time. We use Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests to check whether the time series is stationary or not. KPSS test uses a null hypothesis that an observable time series is stationary against the alternative of a unit root (Kwiatkowski, Phillips, Schmidt, & Shin, 1992).
2. Seasonality: Seasonality can be another component of the pattern in a time series. Seasonality refers to conditions where a seasonal pattern exists in a time series. It is essential to deseasonalise time series when they have a seasonal pattern. We can detect seasonality simply when time series have a large autocorrelation at fixed seasonal lags. We used the STL function in R to decompose time series into seasonal, trend and random components R Core Team (2018). STL is a seasonal trend decomposition procedure based on LOESS that is robust to extreme observations and missing data (Makridakis, Wheelwright, & Hyndman, 2008).
Suppose a time series is defined by the model $Y_t = S_t + T_t + E_t$, where S_t denotes the seasonal component, T_t denotes the trend component, and E_t denotes the errors. (Note that, time series can also have a multiplicative form $Y_t = S_t * T_t * E_t$ which can be expressed in an additive form by taking the natural logarithm of both sides of this model.) We can de-trend a time series by subtracting the trend from series, $X_t = Y_t^* - T_t$, and deseasonalise it by subtracting the seasonality part, $Z_t = X_t - S_t$, and the remainder is error term, $e = Y_t^* - T_t - S_t$. For such a simplified time series model, we can measure trend and seasonality using Eqs. (14) and (15), respectively (Wang et al., 2009).
$$\text{Trend strength} = 1 - \frac{\mathrm{Var}(e_t)}{\mathrm{Var}(Z_t)} \tag{14}$$
$$\text{Seasonality strength} = 1 - \frac{\mathrm{Var}(e_t)}{\mathrm{Var}(X_t)} \tag{15}$$
3. Non-linearity: There are different methods to test for non-linearity. We computed this metric with Terävirta's non-linearity test (Teräsvirta, Lin, & Granger, 1993). This test uses a statistic $T \log\left(\frac{SSE_{nl}}{SSE_l}\right)$ where SSE_{nl} and SSE_l correspond to the sum-of-squares of the non-linear and linear models, respectively.
4. Skewness: Skewness is a metric used to measure the asymmetry of a distribution which shows the level that a distribution differs from a normal distribution. The skewness of a normal distribution is zero. The skewness can be measured with Eq. (16) for univariate time series such that,
$$\text{Skewness} = \frac{1}{n\sigma^3}\sum_{t=1}^{n}(Y_t - \overline{Y})^3, \tag{16}$$
where $\overline{Y}$ is the mean of the time series, σ is the standard deviation of the time series, and n is the number of observations.
The skewness is close to zero if the time series is similar to a normal distribution. It has a negative value for time series that are skewed to the left and takes on a positive value for time series that are skewed to the right.
5. Kurtosis: Kurtosis is a measure that shows whether distribution or data are heavy-tailed or light-tailed in comparison with a normal distribution. Time series that have high kurtosis have heavy tails and time series with low kurtosis have light tails. The kurtosis can be measured with Eq. (17) for a univariate time series,
$$\text{Kurtosis} = \frac{1}{n\sigma^4}\sum_{t=1}^{n}(Y_t - \overline{Y})^4, \tag{17}$$
where $\overline{Y}$ is the mean of time series, σ is the standard deviation of the time series, and n is the number of observations.
6. Spectral Entropy: Spectral entropy is a measure used to quantify the complexity of signals in the non-linearity analysis literature. Entropy is a common measure by which many researchers quantify the predictability of a time series (Garland et al., 2014; Goerg, 2013). We use the *Shanon entropy* which describes the relative contribution of frequencies to the unpredictability. This measure is computed using Eq. (18),

$$\text{entropy} = -\int_{-\pi}^{\pi} \hat{f}(\lambda)\log\left(\hat{f}(\lambda)\right)d\lambda, \tag{18}$$

where $\hat{f}(\lambda)$ is the spectral density of the data. The spectral density describes the power or strength of a time series as a function of frequency λ. Small values of this measure indicate a high signal to noise ratio and unpredictability of the time series.

The above-mentioned features are applied to the raw time series and calculated using the '*tsfeatures*' package in R Hyndman et al. (2019). Moreover, non-linearity, skewness, and kurtosis are applied to the seasonally adjusted series. Since the range of these metrics are different and vary across the series, we transformed them so that they are defined to lie within the [0, 1] range. We use two functions to transfer the results for raw time series and deseasonalised time series as discussed in Wang et al. (2009). We also tested the stationarity of the demand time series. According to the KPSS test, 694 of demand series are stationary. We differenced the non-stationary demand time series at lag one and tested them again to make sure they are all stationary. We calculated the above-mentioned features over the training set for three different groups of SKULs with different levels of volatilities.

Table 1 summarises the six time series featured described above, plus the deseasonlisation of three of them, for the three groups and for all of the demand time series. Since the range of the values is between [0, 1], we can say that the investigated demand time series have a high level of *entropy*, *kurtosis*, and *skewness*; note that these values are 0.95, 0.77, and 0.68, respectively. The series show a low level of *seasonality* and *non-linearity* with values 0.02 and 0.10, respectively. The price-demand relationship is often non-linear, but in our case the non-linearity feature is weak and we can use a linear model (piecewise linear) to forecast the series.

Table 1 along with Fig. 3 show that there is a positive relationship between CoV and entropy of demand series indicating that when the CoVs are increasing the time series is more difficult to forecast due to the volatility in demand series. Moreover, the groups with high CoV values have higher values of kurtosis and skewness that are indicative of the time series volatility and forecastability. Therefore, we rely on the CoV as an important and more common feature of demand series in the supply chain context and it is indicative of demand variability and forecastability.

5. Empirical results and discussion

The data were split into a training-set and a test-set. The first 104 weeks were used as the training-set and to estimate the model parameters. The last eight weeks are considered as the test-set and used to measure the performance and accuracy of the forecasting models. Since price is known in advance, we use it as an explanatory variable and generate forecasts based on a rolling origin for the next eight weeks.

One natural way to reduce uncertainty and control volatility is to combine forecasts from different methods. We use forecast combination as another method to forecast demand. Forecast combination has been advocated by researchers in numerous studies Armstrong (2001). We take a simple average (equal weights) of three different best performing models, ARIMAX, DLR, and SVR. The ARIMA, ETS along with their advanced variations ARIMAX and ETSX are used as benchmark methods. Theta models is also used as another successful benchmark that utilities decomposition-based forecast combination method Makridakis and Hibon (2000).

Table 1
Time series feature and demand volatility

Criteria	A	B	C	All
Deseasonlised non-linearity	0.12	0.18	0.16	0.15
Deseasonlised skewness	0.40	0.82	0.97	0.69
Deseasonlised kurtosis	0.57	0.91	0.99	0.80
Non-linearity	0.14	0.10	0.05	0.10
Seasonality	0.04	0.02	0.01	0.02
Trend	0.45	0.30	0.23	0.34
Skewness	0.38	0.82	0.94	0.68
Kurtosis	0.54	0.88	0.99	0.77
Entropy	0.91	0.97	0.99	0.95

Since the scale of the demand varies across series, we need to use scale-independent criteria to validate the accuracy of models. Mean absolute scaled error (MASE) is used to evaluate the validity of each of the presented models Hyndman (2006). MASE is defined as follows,

$$\text{MASE} = \frac{1}{h} \times \frac{\sum_{t=n+1}^{n+h} |Y_t - \hat{Y}_t|}{\frac{1}{n-1}\sum_{t=2}^{n}\left|Y_t - Y_{t-1}\right|}, \tag{19}$$

where h is the forecasting horizon, n is the sample size, Y_t is the observed value of the series at time t, and $\hat{Y}_t$ is the generated forecast.

MASE values smaller than one means to what extent the generated forecast, on average, is better than the h-step ahead in sample naive forecast. Smaller values indicate a better forecast on average. Table 2 shows the median MASE for the different models for eight-step ahead forecasts.

The results slightly differs across the different groups and methods and can be described as follows. For the series with low volatility, ARIMAX has the lowest MASE followed by SVR. While adding covariates to ETS did not improve its accuracy, it did improve the accuracy of the ARIMA model significantly. ETSX showed the lowest accuracy. This is surprising since ETSX benefits from the inclusion of an explanatory variable. A close look at the results of ETSX revealed that ETSX generates fairly good results during non-promotional periods but dramatically over forecasts promotional sales. As a result, this creates, on average, a large forecasting error. DLR and SVR showed robust performance and higher levels of accuracy when compared with other statistical and ML models across the different groups. While ANN works well for series with low volatility, its accuracy decreases dramatically for series with higher volatility. These results show that this method requires further investigation since it contradicts other results found in the literature Ali et al. (2009, 2001). Defining a new architecture for the ANN, training it across the series, identifying the input and the most relevant type of data with longer history are only some of the ways that may be considered to improve ANN results Crone and Kourentzes (2010). The COM method has the highest accuracy, on average. However, it is outperformed by ARIMAX and HR-ARIMA models for individual groups of series. The superior performance of the COM method is due to combination of ML and statistical methods that cancel out the downside of each other. This is consistent with the results of M4 competition where a combination of ML and statistical methods found to remarkably improve the forecasting accuracy Makridakis, Spiliotis, and Assimakopoulos (2018).

We used pairwise t-tests to determine whether the forecasting accuracy of each model significantly differs when series show different levels of volatility. Table 3 shows the statistically significant differences when a model is used to forecast across the different groups of SKULs with different volatilities. In general, we can see that volatility significantly impacts the performance of the models. The HR-ARIMA and COM models are the only robust models that perform well across the three groups of SKULs with different levels of volatilities. There is no statistically significant difference in the accuracy of models as the CoV increases. The ARIMAX, DLR and SVR models have shown robust performance between the low volatility and moderate volatility groups. ARIMA, ETS, ETSX, ANN, and Theta models are not robust across groups of SKULs with different volatilities.

Fig. 4 shows the performance of the different models with respect to their CoV. In general, the MASE increases as CoV increases. We can see

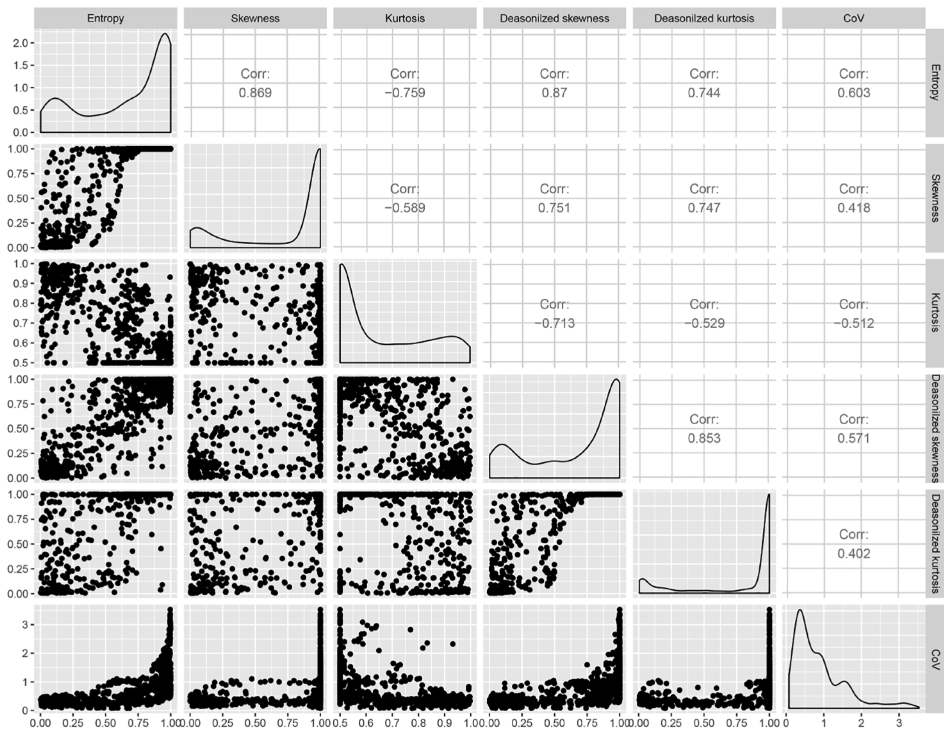

Fig. 3. The relationship between time series characteristics.

Table 2
Eight-step ahead forecasting accuracy as determined by median MASE.

Forecasting Models	SKULs groups			
	A	B	C	Total
ANN	0.201	0.323	0.404	0.293
ARIMA	0.253	0.470	0.71	0.467
ARIMAX	**0.181**	**0.189**	0.250	0.203
COM	0.195	0.198	0.224	**0.188**
DLR	0.207	0.209	0.238	0.215
ETS	0.255	0.573	1.112	0.606
ETSX	0.291	0.648	1.130	0.640
HR-ARIMA	0.221	0.228	**0.219**	0.197
SVR	0.189	0.206	0.230	0.209
Theta	0.258	0.451	0.712	0.454

that DLR, ARIMAX, SVR, HR-ARIMA, and COM models are more robust to CoV variations. While the COM, SVR, and DLR models have clearly better performance than HR-ARIMA for smaller values of CoV, HR-ARIMA outperforms other models for CoV values greater than 1.5. Theta has also shown good performance for series with a CoV smaller than 1, but its accuracy deteriorates fast for more volatile series. The ETS and ETSX models exhibit a similar performance for series with small CoV while ETSX performs poorly for larger values of CoV. ANN is the least accurate model for series with CoV smaller than 1.5 but outperforms ETS and ETSX for larger values of CoV.

In general, of the statistical models considered in this section, the constructed hybrid model has the highest accuracy across moderate and highly volatile demands. This model uses Algorithm 1 and benefits from a combination of features from a piecewise regression and an ARIMA model to estimate promotional and non-promotional demands, respectively. The superior performance of the hybrid model is the result of decomposing sales into different components and using the appropriate model to forecast each of them. This is consistent with the other results found in the literature Pai and Lin (2005). Of the statistical models examined here, the ETSX model has a poor performance and is not recommended to use for forecasting volatile time series during promotions. Further research needs to be undertaken to investigate the factors leading to the poor performance of this model. DLR seems to be promising and SVR generates robust forecasts across different values of CoV. Our results are consistent with the results described by Wang et al. (2009). Surprisingly, the Theta model did not generate robust and accurate forecasts across the different groups and its accuracy decreases as CoV increases. This might be because Theta model is unable to deal with promotions as it does not use explanatory variable.

5.1. Inventory performance

We now turn our attention to the utility of the forecast. We do so by investigating the inventory performance of the forecasting models. In our case, inventory follows the Order Up To (OUT) policy with one week of lead time. Under the OUT policy, in period t the company receives the orders that have been placed in period $t-1$. Demand is satisfied with inventory, if available, and an order is placed according to the forecasted demand and inventory position. We can compute the order quantities and inventory position using Eqs. (20) and (21), respectively,

$$O_t = \widehat{D}_t + SS_t - IP_t, \tag{20}$$

$$IP_t = IP_{t-1} + O_{t-1} - D_t, \tag{21}$$

where O_t is the order at time t, SS_t is the safety stock at time t, and IP_t is the inventory position at time t Johnson and Thompson (1975). We quantify the safety stock quantity by $SS_t = \phi^{-1}(\alpha)\sigma_d$ where α is the targeted service level, $\phi^{-1}(.)$ is the inverse of cumulative distribution of the standard normal probability density function, and σ_d is the standard deviation of the forecast error.

In general, there are three different types of metrics that can be employed to measure the inventory performance of forecasting models: financial, operational, and service. In a supply chain context, operational and service metrics are frequently used since financial data are not often available Petropoulos, Wang, and Disney (2019, 2011).

We use both first-order and second-order functions to measure the inventory performance of the models. The first-order function includes

Table 3
P-values: Volatility impact on the forecasting accuracy (MASE) of models.

Forecasting Models		SKULs groups		
		A	B	C
ANN	A	–	**0.012**	**0.010**
	B	**0.012**	–	**0.004**
	C	**0.010**	**0.004**	–
ARIMA	A	–	**0.021**	**0.002**
	B	**0.021**	–	**0.000**
	C	**0.002**	**0.000**	–
ARIMAX	A	–	0.145	0.051
	B	0.145	–	**0.001**
	C	0.051	**0.001**	–
COM	A	–	0.010	0.124
	B	0.010	–	0.067
	C	0.124	0.067	–
DLR	A	–	0.158	**0.012**
	B	0.158	–	**0.010**
	C	**0.012**	**0.010**	–
ETS	A	–	**0.005**	**0.000**
	B	**0.005**	–	**0.000**
	C	**0.000**	**0.000**	–
ETSX	A	–	**0.001**	**0.000**
	B	**0.001**	–	**0.003**
	C	**0.000**	**0.003**	–
HR-ARIMA	A	–	0.126	0.170
	B	0.126	–	0.104
	C	0.170	0.104	–
SVR	A	–	0.072	**0.001**
	B	0.072	–	**0.042**
	C	**0.001**	**0.042**	–
Theta	A	–	**0.013**	**0.004**
	B	**0.013**	–	**0.000**
	C	**0.004**	**0.000**	–

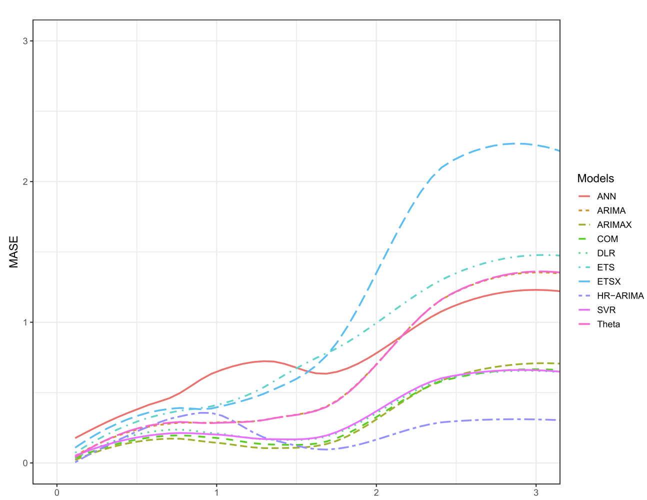

Fig. 4. MASE with respect to CoV.

the total inventory costs and is a summation of the holding costs (h) and backlog costs (b). Here h and b together define the targeted service level $\alpha = \frac{b}{h+b}$. Holding cost is imposed when the inventory at time t is higher than the demand at the same period, and backlog cost is imposed when the demand at time t is higher than the inventory at the same period. Second-order functions include the order and inventory variance and are computed empirically.

Table 4
Inventory performance (α = 0.99).

Models	Criteria				
	CoV	Inventory cost	Order variance	Inventory variance	Fill rate
ANN	A	5.289	2.239	2.123	0.983
	B	5.610	3.562	3.131	0.964
	C	6.854	5.602	6.295	0.975
	All	17.754	5.845	5.317	0.976
ARIMA	A	5.774	2.435	2.152	0.975
	B	6.886	5.203	2.501	0.963
	C	7.930	6.331	4.338	0.911
	All	20.591	6.112	5.988	0.937
ARIMAX	A	3.919	1.700	2.168	0.986
	B	4.859	5.203	3.080	0.988
	C	6.245	6.771	5.706	0.994
	All	**15.024**	5.343	5.543	0.982
COM	A	5.555	2.164	2.230	0.985
	B	6.744	4.512	3.405	0.992
	C	8.485	6.211	6.939	0.997
	All	20.785	**4.805**	5.424	0.994
DLM	A	5.109	2.131	2.216	0.975
	B	6.196	4.515	4.404	0.968
	C	6.929	5.139	8.874	0.987
	All	18.235	5.925	5.482	0.984
ETS	A	5.527	2.351	2.095	0.973
	B	6.535	4.589	3.651	0.967
	C	7.933	7.052	6.466	0.937
	All	19.996	6.183	6.654	0.955
ETSX	A	5.183	2.226	2.262	0.990
	B	5.614	3.723	3.603	0.963
	C	10.543	7.375	7.434	0.964
	All	21.341	6.227	5.987	0.975
HR-ARIMA	A	5.889	2.436	5.498	0.971
	B	5.838	4.116	4.519	0.982
	C	5.938	4.726	4.627	0.994
	All	17.666	4.828	5.811	0.985
SVR	A	5.517	1.944	2.300	**0.997**
	B	7.224	5.012	3.465	0.965
	C	8.393	6.186	7.667	0.993
	All	21.135	6.813	6.167	0.986
Theta	A	5.474	2.559	2.164	0.992
	B	6.565	6.627	3.488	0.965
	C	6.602	4.453	5.491	0.976
	All	18.642	6.282	**5.020**	0.986

In order to be able to directly compare and summarise the variances of the orders and the inventory, we standardise the data by *minmax* method where we subtract the min of the series and divide it by the range (max–min) of the corresponding series. We filtered our inventory analysis only to studying the 694 stationery series. The first 104 periods are treated as the in-sample data and last eight periods as out-sample data. The initial value of orders for promotional and non-promotional periods are set equal to the mean of in-sample similar promotional sales and the mean of non-promotional sales, respectively. The initial value of inventory is set equal to the mean of the forecasted sales plus safety stock, minus the mean sales. We then calculate the inventory and orders at each period using Eqs. (20) and (21). Since we do not have a sales forecast for period 113, the orders are not calculated for period 112. Therefore, the order variance is calculated based on the seven periods (periods 104–111).

Tables 4 and 5 show the inventory performance of the different

Table 5
Inventory performance ($\alpha = 0.95$).

Models	CoV	Inventory cost	Order variance	Inventory variance	Fill rate
ANN	A	0.949	2.050	1.868	0.894
	B	0.935	2.861	1.178	0.953
	C	1.055	3.854	3.929	0.956
	All	2.939	2.845	5.317	0.896
ARIMA	A	1.014	2.178	1.942	0.894
	B	1.116	3.864	1.699	0.869
	C	1.140	3.565	2.320	0.877
	All	3.271	3.112	1.988	0.898
ARIMAX	A	0.698	1.554	1.957	0.957
	B	0.811	4.450	2.199	0.969
	C	0.973	4.617	3.615	0.946
	All	**2.483**	3.343	2.543	0.945
COM	A	0.973	1.892	1.978	0.948
	B	1.094	3.228	2.361	0.965
	C	1.248	3.503	4.695	0.970
	All	3.315	**2.805**	2.924	**0.965**
DLM	A	0.912	1.955	1,970	0.945
	B	1.044	3.760	3.008	0.966
	C	1.051	3.358	5.866	0.958
	All	3.008	2.925	3.482	0.959
ETS	A	0.973	2.116	1.849	0.898
	B	1.066	3.352	2.421	0.965
	C	1.183	4.374	3.849	0.947
	All	3.223	3.183	2.654	0.899
ETSX	A	0.921	2.018	2.008	0.941
	B	0.920	2.978	2.435	0.964
	C	1.142	5.032	4.755	0.964
	All	2.985	3.227	2.987	0.936
HR-ARIMA	A	1.018	2.116	5.202	0.918
	B	0.958	2.116	0.939	0.950
	C	0.998	3.276	1.129	0.952
	All	2.975	3.128	7.811	0.942
SVR	A	0.956	1.679	2.044	0.937
	B	1.170	3.497	2.417	0.968
	C	1.236	3.544	5.355	0.950
	All	3.363	2.813	3.167	0.945
Theta	A	0.973	2.289	1.935	0.935
	B	1.099	5.082	2.211	0.967
	C	0.993	2.818	2.910	0.972
	All	3.065	3.282	**2.320**	0.939

forecasting models for $\alpha = 0.99$ and $\alpha = 0.95$, respectively. We have measured the inventory costs so that it include holding and backlog costs, inventory variance, order variance, and fill rate for the different groups of SKULs with different levels of volatilities. In general, inventory and order variance is higher for SKULs with higher volatilities. Inventory costs also increase for more volatile series. One possible reason is that when a series is volatile, it is more difficult to forecast and there is a greater risk of over-forecasting or under-forecast when forecasting a volatile series. As a result, this causes large inventory costs. Although we find that the ARIMAX model has the lowest inventory costs followed by the hybrid model. Interestingly, the inventory costs do not fluctuate much for the HR-ARIMA model across the different groups. The COM approach has the smallest order variance and Theta has the smallest inventory variance. These models both use the combination method and such findings are consistent with the literature Petropoulos et al. (2019). Forecast combination has been shown to reduce the forecast error, and consequently, reduces the inventory variance, and safety stock Hibon and Evgeniou (2005). Finally, models have different fill rates across the different groups of SKULs. For example, COM approach has the highest fill rate on average while the HR-ARIMA model has the highest fill rate for group C. The ETS and ARIMA models have small fill rates for a volatile series in group C.

These results are slightly different when $\alpha^* = 0.95$. ARIMAX has the lowest inventory costs followed by the hybrid HR-ARIMA model. The COM method has the smallest order variance and the second smallest inventory variance after Theta. Fill rates are impacted by the level of volatilities. For example, the ETSX, HR-ARIMA, and Theta models have higher fill rates than the targeted service level which leads to larger inventory costs. Nevertheless, all presented models, except ARIMA and ETS, have shown high fill rates.

6. Conclusion

Forecasting retail sales are of importance for many managerial decisions at different levels of the supply chain. In the modern competitive market, many internal and external factors impact demand in different ways and make them volatile and unpredictable. Promotion is one of the factors that can have differing effects on demand dynamics and can make demand volatile not only over the promotional periods but also over the entire demand series. Volatile demand reduces the supply chain performance and imposes unnecessary costs to supply chain in inventory, supply, and transportation. There is a need for simple, yet accurate models to forecast volatile demand since sophisticated models are not commonly used in practice due to the lack of expertise and resources (Abolghasemi, Eshragh, Hurley, & Fahimnia, 2019a; Hughes, 2001; Makridakis et al., 2018).

We investigated important features of demand such as seasonality, trend, entropy, skewness, kurtosis and showed that the CoV can be used as an important, yet simple and informative statistic to measure the relative volatility of demand time series that are impacted by promotion. We investigated the behavior of 694 SKULs that are impacted by promotions and have different levels of volatilities. We categorised demand series into three groups based on their CoV low volatile, moderately volatile, or highly volatile. We showed that volatility of demands can significantly impact their forecasting accuracy.

We proposed a hybrid model where we decomposed demand time series into baseline demand and promotional demand (uplifts because of promotion). We then constructed a piecewise regression model composing of a model that effectively models demand uplifts during promotional periods, and an ARIMA model to forecast baseline demand. We showed that our proposed hybrid model is simple yet accurate when demand series are highly volatile. This model depicts a robust performance across a series with different levels of volatility. We also evaluated the different types of well-established models in the literature. These models were ARIMA, ARIMAX, ETS, ETSX, DLR, Theta, forecast combination, and common ML algorithms, ANN and SVR. We empirically evaluated their performance and robustness when demand series exhibit different levels of volatility.

We showed that of the models studied presented models, ARIMAX has a superior performance for low and moderately volatile products. Surprisingly, adding a covariate to the ETS model brings about no improvement in accuracy. DLR and SVR showed a similar and robust performance across the different values of CoV. ANN generates better results for a low volatile demand series but works poorly for highly volatile demand series. We also, showed that simple statistical models such as ETS can outperform some of the more sophisticated ML and statistical models when a series is highly volatile. We empirically analysed the inventory performance of the forecasting models. By undertaking a comprehensive analysis of inventory costs, inventory variance, order variance, and fill rate when series have different levels of volatility. Our analysis revealed that the forecast combination effectively reduces the inventory and order variance. We also showed that our proposed hybrid model has a robust performance and small inventory costs. Our models can be used to forecast and improve the required inventory performance when a demand series is highly volatile.

For future research, there is a lot that requires the attention of researchers and practitioners. As we have shown in this paper, the effectiveness of a modelling strategy changes significantly under different levels of volatilities. One can analytically or empirically analyse why some models work better than others in certain conditions. An integrated approach that combines different techniques for promotional and non-promotional periods, and encompass all aspects of the relevant demand features, is a desirable direction for future research. There are many different variables that govern the dynamics of demand, but it could be expensive while also being technically and practically and complicated to consider all of them. Moreover, data may not be available to use. In this paper, price is used as the key variable since there is a strong correlation between price and demand. It is also worth noting that price explains the majority of variation in demand. Considering other influential factors such as other promotional information, special events, holidays and display types and using a multivariate modelling approach to capture the volatility of demand would be a worthwhile avenue of investigation. Another interesting research direction is to analyse different inventory policies and evaluate their performance when demands have different levels of volatility. It would be interesting to identify a CoV threshold to switch from one inventory model to another. Finally, the study focuses on an FMCG company in the Australia's market, this may limit the implications of our findings when applied to other data in different industries. Therefore, the research models may be examined and validated in a different context with different types of data set. It would be beneficial to test different ML and heuristic forecasting methods that estimate the models parameters precisely against the limited amount of available historical data for promotions.

CRediT authorship contribution statement

Mahdi Abolghasemi: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Eric Beh:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Garth Tarr:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Richard Gerlach:** Conceptualization, Methodology, Validation, Investigation, Resources, Supervision, Project administration, Funding acquisition.

Acknowledgements

This study was funded by the Australian Research Council (Grant ID: IC140100032). The authors are grateful to Prof George Athanasopoulos and Prof Aris Syntetos for their invaluable feedback and constructive comments.

References

Abolghasemi, M., Eshragh, A., Hurley, J., & Fahimnia, B. (2019a). Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. arXiv preprint arXiv:1909.02716.

Abolghasemi, M., Hyndman, R.J., Tarr, G., & Bergmeir, C. (2019b). Machine learning applications in time series hierarchical forecasting. arXiv preprint arXiv:1912.00370.

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing, 7*(1), 136–144.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews, 29*(5–6), 594–621.

Ali, Ö. G., Sayın, S., Van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications, 36*(10), 12340–12348.

Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services, 8*(3), 147–156.

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners,, vol. 30*. Springer Science & Business Media.

Armstrong, J. S. & Green, K. C. (2017). Demand forecasting II: Evidence-based methods and checklists.

Assimakopoulos, V., & Nikolopoulos, K. (2000). The theta model: A decomposition approach to forecasting. *International Journal of Forecasting, 16*(4), 521–530.

Aye, G. C., Balcilar, M., Gupta, R., & Majumdar, A. (2015). Forecasting aggregate retail sales: The case of South Africa. *International Journal of Production Economics, 160*, 66–79.

Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews, 11*(10), 203–224.

Blattberg, R. C., & Neslin, S. A. (1993). Sales promotion models. *Handbooks in Operations Research and Management Science, 5*, 553–609.

Brown, R. G. (1959). *Statistical forecasting for inventory control.* McGraw/Hill.

Cachon, G. P. (1999). Managing supply chain demand variability with scheduled ordering policies. *Management Science, 45*(6), 843–856.

Chan, K. -S. & Ripley, B. (2018). TSA: Time Series Analysis. R package version 1.2.

Christopher, M., & Holweg, M. (2011). "Supply chain 2.0": Managing supply chains in the era of turbulence. *International Journal of Physical Distribution & Logistics Management, 41*(1), 63–82.

Christopher, M., & Holweg, M. (2017). Supply chain 2.0 revisited: a framework for managing volatility-induced risk in the supply chain. *International Journal of Physical Distribution & Logistics Management, 47*(1), 2–17.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559–583.

Crone, S. F., & Kourentzes, N. (2010). Feature selection for time series prediction–a combined filter and wrapper approach for neural networks. *Neurocomputing, 73*(10–12), 1923–1936.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., & Leisch, M. F. (2006). The e1071 package. Misc Functions of Department of Statistics (e1071), TU Wien.

Divakar, S., Ratchford, B. T., & Shankar, V. (2005). Practice prize article—chan4cast: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Marketing Science, 24*(3), 334–350.

Donohue, K. L. (2000). Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science, 46*(11), 1397–1411.

Fildes, R., Nikolopoulos, K., Crone, S. F., & Syntetos, A. A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society, 59*(9), 1150–1172.

Garland, J., James, R., & Bradley, E. (2014). Model-free quantification of time-series predictability. *Physical Review E, 90*(5), 052910.

Gilliland, M. (2010). *The business forecasting deal: exposing myths, eliminating bad practices, providing practical solutions,, vol. 27*. John Wiley & Sons.

Goerg, G. (2013). Forecastable component analysis. *International Conference on Machine Learning* (pp. 64–72). .

Günther, F., & Fritsch, S. (2010). neuralnet: Training of neural networks. *The R Journal, 2*(1), 30–38.

Guo, F., Diao, J., Zhao, Q., Wang, D., & Sun, Q. (2017). A double-level combination approach for demand forecasting of repairable airplane spare parts based on turnover data. *Computers & Industrial Engineering, 110*, 92–108.

Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting, 21*(1), 15–24.

Hope, J., & Fraser, R. (2003). *Beyond budgeting: How managers can break free from the annual performance trap.* Harvard Business Press.

Huang, M.-G., Chang, P.-L., & Chou, Y.-C. (2008). Demand forecasting and smoothing capacity planning for products with high random demand volatility. *International Journal of Production Research, 46*(12), 3223–3239.

Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting fmcg retail product sales and the variable selection problem. *European Journal of Operational Research, 237*(2), 738–748.

Huang, T., Fildes, R., & Soopramanien, D. (2019). Forecasting retailer product sales in the presence of structural change. *European Journal of Operational Research*.

Hughes, M. (2001). Forecasting practice: Organisational issues. *Journal of the Operational Research Society, 52*(2), 143–149.

Hyndman, R. (2018). fpp2: Data for Forecasting: Principles and Practice (2nd Edition). R package version 2.3.

Hyndman, R., Kang, Y., Montero-Manso, P., Talagala, T., Wang, E., Yang, Y., & O'Hara-Wild, M. (2019). tsfeatures: Time series feature extraction. *R package version, 1*, 1.

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach.* Springer Science & Business Media.

Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting, 4*(4), 43–46.

Hyndman, R. J. & Athanasopoulos, G. (2014). Forecasting: principles and practice. OTexts.

Johnson, G. D., & Thompson, H. (1975). Optimality of myopic inventory policies for certain dependent demand processes. *Management Science, 21*(11), 1303–1307.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics, 54*(1–3), 159–178.

Ma, S., & Fildes, R. (2017). A retail store sku promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research, 260*(2), 680–692.

Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of sku retail sales forecasting with intra-and inter-category promotional

information. *European Journal of Operational Research, 249*(1), 245–257.
Makridakis, S., & Hibon, M. (2000). The M3-competition: results, conclusions and implications. *International Journal of Forecasting, 16*(4), 451–476.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The m4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting, 34*(4), 802–808.
Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS One, 13*(3), e0194889.
Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications.* John Wiley & Sons.
Meade, N. (2000). Evidence for the selection of forecasting methods. *Journal of Forecasting, 19*(6), 515–535.
Meindl, P., & Chopra, S. (2001). *Supply chain management: Strategy, planning, and operation.* Prentice Hall.
Min, H. (2010). Artificial intelligence in supply chain management: Theory and applications. *International Journal of Logistics: Research and Applications, 13*(1), 13–39.
Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine, 22*(19), 3055–3071.
Narayanan, A., Sahin, F., & Robinson, E. P. (2019). Demand and order-fulfillment planning: The impact of point-of-sale data, retailer orders and distribution center orders on forecast accuracy. *Journal of Operations Management, 65*(5), 468–486.
Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega, 33*(6), 497–505.
Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic linear models.* Springer.
Petropoulos, F., Wang, X., & Disney, S. M. (2019). The inventory performance of forecasting methods: Evidence from the m3 competition data. *International Journal of Forecasting, 35*(1), 251–265.
Priore, P., Ponte, B., Rosillo, R., & de la Fuente, D. (2018). Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments. *International Journal of Production Research, 57*(11), 3663–3677.
R Core Team (2018). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.
Robert C. Blattberg, R. B. & J, E. (1995). How Promotions Work. Marketing Science, 14(3):122–132.
Scholz-Reiter, B., Heger, J., Meinecke, C., & Bergmann, J. (2012). Integration of demand forecasts in abc-xyz analysis: Practical investigation at an industrial company. *International Journal of Productivity and Performance Management, 61*(4), 445–451.
Serdarasan, S. (2013). A review of supply chain complexity drivers. *Computers & Industrial Engineering, 66*(3), 533–540.
Shah, C. (1997). Model selection in univariate time series forecasting using discriminant analysis. *International Journal of Forecasting, 13*(4), 489–500.
Silver, E. A., Pyke, D. F., Peterson, R., et al. (1998). *Inventory management and production planning and scheduling, vol. 3.* New York: Wiley.
Strijbosch, L. W., Syntetos, A. A., Boylan, J. E., & Janssen, E. (2011). On the interaction between forecasting and stock control: The case of non-stationary demand. *International Journal of Production Economics, 133*(1), 470–480.
Svetunkov, I. (2018). smooth: Forecasting using state space models. R package version 2.4.5.
Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*(1), 1–26.
Syntetos, A. A., Boylan, J. E., & Croston, J. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society, 56*(5), 495–503.
Teräsvirta, T., Lin, C.-F., & Granger, C. W. (1993). Power of the neural network linearity test. *Journal of Time Series Analysis, 14*(2), 209–220.
Torkul, O., Yılmaz, R., Selvi, I., & Cesur, M. R. (2016). A real-time inventory model to manage variance of demand for decreasing inventory holding cost. *Computers & Industrial Engineering, 102*, 435–439.
Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the operational Research Society, 66*(2), 299–307.
Van Heerde, H. J., Leeflang, P. S., & Wittink, D. R. (2002). How promotions work: SCAN*PRO-based evolutionary model building. *Schmalenbach Business Review, 54*(3), 198–220.
Villegas, M. A., Pedregal, D. J., & Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. *Computers & Industrial Engineering, 121*, 1–7.
Wang, X., & Petropoulos, F. (2016). To select or to combine? The inventory performance of model and expert forecasts. *International Journal of Production Research, 54*(17), 5271–5282.
Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery, 13*(3), 335–364.
Wang, X., Smith-Miles, K., & Hyndman, R. (2009). Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. *Neurocomputing, 72*(10–12), 2581–2594.
Zeileis, A. (2019). dynlm: Dynamic Linear Regression. R package version 0.3-6.
Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.