From predictive to prescriptive analytics: A data-driven multi-item newsvendor model

Sushil Punia*, Surya Prakash Singh, Jitendra K. Madaan

Department of Management Studies, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

A R T I C L E   I N F O

A B S T R A C T

This paper considers a multi-item newsvendor problem with a capacity constraint (Z). The problem has already been addressed in the literature using the classical newsvendor problem. However, provided solutions made assumptions for demand distributions, which are often incorrect and led to errors in the inventory optimization. This research proposes a distribution-free and completely data-driven solution approach to Z. The proposed approach uses sample demand data as input, and machine (and deep) learning methods with empirical risk minimization principle to find order quantities. A heuristic is developed using hierarchies of the retail products to perform multi-item inventory optimization when a capacity constraint is active. The proposed approach is tested on a real-world dataset of retail products. The results from the proposed method are compared with data-driven max-min and empirical inventory optimization methods, and it outperformed all of them. The machine (and deep) learning-based demand forecasting methods (part of the proposed approach) providing better results than neural networks, multiple regression, arima, arimax, etc. Finally, a comparison of total inventory cost from the proposed, max-min, and empirical inventory optimization methods are carried out, and it is observed that the proposed data-driven approach leads to a significant reduction in inventory cost.

1. Introduction

Many firms, including manufacturers, distributors, and retailers, commonly face two critical problems. The first problem is to decide the production or order quantity for products with short life-cycle before selling season. The second problem for firms is to do the allocation of a limited resource. Under the first problem, if the order quantity for a product is higher than its demand, then the retailer will incur obsolescence cost; and if the order quantity for a product is lower than the demand, then the retailer will have opportunity and goodwill loss. Thus, a trade-off between ordering too much or too little is a crucial decision to make for a retailer. A newsvendor model, and its variations very well represent this inventory management problem. This type of problem is relevant to many industries, such as fast-food, fashion clothing, groceries, bakeries, sports goods, and perishable products industry.

In a classic newsvendor problem, a retailer minimizes the expected total cost of ordering items from an upstream supply chain partner. To find out the order quantities for products, retailer assumes a specific distribution of the stochastic demand, and that leads to errors in order quantities, and consequently increase in cost. Although a large number of retailers use these theoretical newsvendor models, the assumption of

a specific statistical distribution for demand does not hold well in reality. Moreover, the pattern of demand changes with the business environment, thus changes the distribution of demand over time, too [26]. Such an approach that is free from the assumption of demand distribution shall be useful for inventory management in practice.

Under the second problem, firms have to do allocations of a limited resource (e.g., budget, shelf-space, or time, etc.) among multiple products/items. This problem is known as a resource allocation problem. For example, a retailer has to allocate limited shelf-space to different variants of a particular food item, or similarly, a fashion store has limited space for different colors and sizes of the same garment.

The paper address these two problems by formulating a multi-item inventory optimization problem with a capacity constraint and providing a data-driven solution to it. Recently, some efforts were made to develop data-driven solutions for the newsvendor problem [22]. However, literature is somewhat fragmented, and multi-item case with a capacity constraint is not addressed. Therefore, this paper address this gap through a data-based solution. A large real-world dataset consisting of demand and business environment data is used as input, and advanced machine (deep) learning algorithms are used to solve the model to obtain order quantities for products. The proposed framework and subsequent contributions and insights are elaborated herewith.

* Corresponding author.
E-mail addresses: sushil.punia@dms.iitd.ac.in (S. Punia), jmadaan@dms.iitd.ac.in (J.K. Madaan).

1.1. Contributions and insights

1.1.1. A machine (deep) learning-based demand estimation method

As a part of the newsvendor model, it is needed to estimate the products' demand for future periods. So, first, we tried to find out an accurate demand estimation method that can be used in the proposed data-driven inventory optimization model. The advanced machine learning method (e.g., random forest, feed-forward neural network, etc.), and deep learning method (e.g., deep neural network) are used to estimate the demand for future periods. The predictions from these methods are compared with widely used time-series and regression methods to investigate the performance of the proposed method.

1.1.2. A data-driven inventory optimization method

To overcome the limitations of a classical newsvendor inventory model, we proposed a novel quantile-regression and machine learning-based approach to find the optimal order quantities for products. The proposed optimization approach integrates the demand estimation and inventory optimization steps eliminating redundant efforts to perform two separate steps. The maximal approximation [16,35], and normal distribution based empirical methods are used to adjudge the performance of the proposed method. In data-driven models, Gallego and Moon [16] provided a method to solve a multi-item inventory optimization under a budget constraint. Because it is a simple yet effective data-driven model, it is used as a benchmark to test our solution.

1.1.3. A heuristic for multi-item inventory optimization with a limited capacity

A heuristic is proposed to determine the optimal order quantities for multiple products under a capacity constraint. In this novel heuristic, top-down hierarchies of the products are used along with demand data to provide the complete data-driven solution to the multi-item inventory optimization with a limited capacity problem. Different possible cases of the problem are discussed, and solutions are provided for the same.

The goal of inventory management is to decide order quantities to procure products from an upstream supply chain member at minimum cost. So, we calculated the total inventory cost for all products, and the impact of different demand estimations and inventory optimization techniques are analyzed on the total inventory cost.

1.2. Organization of the paper

The remainder of this paper is organized as follows. In Section 2, an overview of related literature is discussed, and research gaps are identified. The problem description and proposed data-driven models for inventory optimization are presented in Section 3. Section 4 contains detailed empirical analysis, results, and discussions. Finally, Section 5 contains the conclusion and directions for future work.

2. Related literature

The newsvendor models (hereafter, NVM) in operations management is a widely discussed inventory management problem. To tackle the stochastic nature of demand in NVM, several approaches have been suggested to model the demand uncertainty. The most common approach, in research and practice both, is to assume a specific distribution for demand and calculate its parameters; and, afterward, use the parameters to solve the NVM to find optimal order quantities. The approach is available in several textbooks [37] and review papers [33]. This approach is included in our analysis to benchmark the results of the proposed model [22].

Another class of methods is data-driven models. The first approach under data-driven models is robust optimization, i.e., the optimization of NVM is performed using partially available demand information [9]. Scarf [35] presented the NVM in which only the mean and variance of the demand data is known. Using a mathematical proof, he provided an ordering rule to find the order quantity with the worst and best possible distributions with given mean and variance. This approach is known as a min-max approach. Gallego and Moon [16] extended this idea in several directions underlining its practical value under the conservative approach of using only mean and variance. It is extended to different cases, where second buying opportunity is available, where fixed ordering cost is involved, to random yield, and a multi-item under a capacity constraint. Random yield is a case where yield from production is assumed to be a random variable.

Bertsimas and Thiele [7] extended the research on robust optimization by proposing a deterministic, numerically tractable methodology. In their model, demands are considered with heterogenous demand distributions overtime on echelons. The authors claimed that under robust optimization (knowing mean and variance of demand only), their approach outperforms the models assuming some specific distribution of the demand variable. Another approach under data-driven models is sample average approximation, where a sample of the demand data is used to solve the NVM [27]. Levi et al. [26] provided the distribution-specific tighter bounds version of Levi et al. [27] for single period featureless NVM and showed that sample average approximation could perform at par with the empirical model. Huber et al. [22] applied the sample average approximation to NVM with multi-feature data and found it is performing better than empirical methods in some of the cases.

In data-driven methods, a more recent approach to use demand data for solving NVM advocates integrating the demand estimation and optimization steps into a single step. Ban and Rudin [5] proposed a big data newsvendor problem in which the authors used the machine learning algorithms along with empirical risk minimization and kernel weights optimization to solve a multi-feature NVM. The authors showed that the empirical risk minimization approach is equivalent to a high-dimensional quantile regression and can be solved by convex optimization methods. Oroojlooy, Snyder, and Takáč [30] used the deep neural networks to solve the NVM using an integrated method to solve NVM and extended the approach for (r, Q) policy. However, they have not used quantile regression for the same. A similar approach is proposed by Huber et al. [22], in which the authors used the non-linear machine learning methods like a decision tree and single-layer neural network in quantile regression to solve the single item NVM.

2.1. Time-series, machine and deep learning methods for demand forecasting

The commonly used methods for demand forecasting are time-series methods such as exponential smoothing method (Hyndman et al. 2002; Gardner Jr. 2006), Auto-Regressive Integrated Moving Average (ARIMA) (Chu & Zhang, 2003; Makridakis & Hibon, 1997), Kalman filters (Choi et al., 2011; Xie, Song, Sirbu, & Wang, 1997) and so forth that mainly try to model the trend and cyclicity in a demand series. The use of some multivariate time-series methods such as ARIMAX is also reported in the literature. However, artificial intelligence methods are preferred for handling covariates forecasting problems where data on external business variables are present [31]. The covariates methods establish relationships among demand and independent variables. Machine learning (ML) methods gained popularity in recent years for demand forecasting. The ML methods model patterns in the demand based on correlations with independent variables. In ML methods, artificial neural networks (ANN) are widely used for demand prediction and reported improvements in the accuracy of forecasts in various cases [1,3].

Random forest (RF) is based on regression trees and bagging algorithms. The advantages of RF over NNs are in terms of interpretability and accuracy on non-time-series data [14]. The latest addition to a set of machine learning methods is deep learning [32]. The deep neural network, recurrent neural networks, etc. were used for forecasting in

several domains. As reported by Loureiro, Miguéis, and da Silva [28], results were encouraging to explore further the use of deep learning for sales forecasting.

2.2. Research gaps and main contribution

Within the line of research for multi-feature data-driven NVM optimization, some research gaps exist that we identified and aim to bridge through our work. The first research gap is that random forest, a widely used machine learning algorithm for demand forecasting in retail [14], and similarly, deep neural networks are not used to model a priori non-linear multi-feature relationships between order quantities and demand data and their impact on inventory performance. The second research gap is that multi-item inventory optimization under the capacity constraint case is not discussed in data-driven decision models for multi-feature NVM. The paper aim to bridge these gaps in the literature by investigating how does a retailer shall use sales data to optimize its inventories under limited resources? To address the research question, paper propose to design a data-driven solution to a multi-item inventory optimization under a capacity constraint. The primary contribution to research in comparison with data-driven approaches is summarized in Table 1.

3. Methodology

In this section, we present the problem description and proposed methodology for determining the optimal order quantities for products in multi-item inventory management. This section consists of three parts: 1) demand estimation techniques, 2) inventory optimization techniques, and 3) multi-item inventory optimization under capacity constraint. Following are the main list of notations used in the paper to explain the proposed methodology:

- d_i = random variable for the demand for an item i
- c_i = purchasing cost for each unit of item i
- p_i = selling price for each unit of item i
- h^i = overage/holding cost per each unit of item i
- b^i = underage/backordering cost for each unit of item i
- m = total no. of items/products
- G_n = total no. of items in a product group
- C = total resource capacity for all products of a product group
- $Q_i or Q^i$ = order quantity for item i
- $\pi_i(Q_i)$ = total cost of inventory for item i
- Q_i^u = unconstrained order quantity for item i
- $\alpha^i = b^i/(b^i + h^i)$, the critical fractile (optimal service level) for unconstrained newsvendor item i
- $\beta^i = v^i/(b^i + h^i)$, is the ratio of the capacity requirement to the marginal revenue for item i

The other notations will be defined as and when required.

3.1. Problem description

A multi-item newsvendor problem with unknown demand distributions is considered. The inventory system has a single linear constraint, which represents a limited capacity or a limited budget available with a retailer to spend on the items. The retailer has to decide the optimal quantities required to be ordered before the selling season. As this is variant of newsvendor problem (hereafter, NVP), so as in a classical NVP, overage/holding cost of h, and underage/backordering cost of b for each unit is considered. In a (single item) newsvendor model, the objective function is to minimize the total expected cost, which is given by:

$$\min_{q^i} E[b^i(d^i - Q^i)^+ + h^i(Q^i - d^i)^+], \tag{1}$$

Table 1
Comparison of data-driven inventory optimization approaches.

Paper	Data used	Demand Forecasting Technique			Inventory optimization		Additional remarks
		TS	ML	DL	Multi-item*	Constraints	
1) Scarf [35]	Mean and variance	–	–	–	–	–	Proposed Min-max approach
2) Gallego & Moon [16]	Mean and variance	–	–	–	✓	Capacity	Extended version of (1)
3) Bertsimas & Thiele [7]	Mean and variance	–	–	–	–	–	Used heterogenous demand distributions over the time on echelons
4) Levi et al. [26,27]	Demand data	–	✓	–	–	–	Used weighted mean spread and featureless data
5) Ban & Rudin [5]	Demand data	–	✓	✓	–	–	Used multi-feature data and quantile regression (linear) for optimization
6) Oroojlooy et al. [30]	Demand data	–	–	✓	–	–	Used revised loss function for optimization
7) Huber et al. [22]	Demand data	✓	✓	–	–	–	Used multi-feature data and quantile regression (non-linear) for optimization
Proposed	Demand data	✓	✓	✓	✓	Capacity	Using multi-feature data, quantile regression(non-linear) optimization, deep learning, cross-sectional hierarchies, multi-product, capacity constraint, and a heuristic.

TS – time-series; ML – Machine Learning; DL- Deep Learning; and * - multi-item solutions shall valid only when some constraint is active on them.

where Q_i is the order quantity, and d_i is the random demand for an item i. If $F^i(.)$ is the cumulative density function of the demand distribution of item i and $F^{i-1}(.)$ is its inverse. The optimal solution to this (unconstrained) problem is given by:

$$Q^{i*} = F^{i-1}\left(\frac{b^i}{b^i + h^i}\right) = F^{i-1}(\alpha^i), \tag{2}$$

In a case of multiple items with no capacity constraint, the optimal order quantity for each item can be found out by using Eq. (2) for each item. Moreover, the obtained order quantities will also be optimal when the available capacity is greater than or equal to the sum of ordered quantities.

The actual demand distributions for items, $F^i \ \forall \ i = 1, 2, ..., m$, are unknown. Only the historical demand data, $S = \{(d_j^1, X_j^1), ..., (d_j^i, X_j^i), ..., (d_j^m, X_j^m)\}_{j=1}^n$, where d_j is the demand and X_j is a vector of its covariates with independent features (e.g., price, no. of clicks, promotion, etc.) for each item (i) are known. In this case, either we have to assume the distribution for the demand, or we must determine the demand distribution from the historical demand data to apply inventory optimization. Both methods have their limitations, which are already discussed and well documented in the literature also, e.g. [20]. The paper present both assumption-based (empirical) and data-based methods for inventory optimization.

Overall, the following sections present the proposed solution methodology consisting of demand estimation, and multi-item inventory optimization steps to develop a complete data-driven model to find the solution to the above-described problem. In addition, appropriate inventory optimization methods are also discussed to make the discussions complete.

3.2. Demand estimation through machine/deep learning

As the distributions of the actual demand for the products are not known, so general forecasting methods are used for demand estimation. As explained in Section 2, ML methods such as random forest (RF) and deep neural network (DNNs) are used to estimate the demand.

Random Forest (RF) algorithm uses the principle of *bagging* to reach the final demand prediction. Initially, a q number of bootstrap samples of size, S_n are randomly selected from the demand data with equal probability and replacement. Say, Φ_q is an i.i.d. random variable. The classification and regression trees (CART) algorithm is applied to each sample $(S_n^{\Phi 1}, ..., S_n^{\Phi q})$ to develop q predicting trees $(\hat{h}(X, S_n^{\Phi 1}), ..., \hat{h}(X, S_n^{\Phi q}))$ [8]. Then demand predictions from the trees are aggregated using Eq. (3) to get the final demand prediction.

$$Y' = \frac{1}{q} \sum_{l=1}^{q} \hat{h}(X, S_n^{\Phi_l}) \tag{3}$$

A deep neural network (hereafter, DNN) takes the demand data and independent variables in the input layer, and then process data using linear and non-linear functions to obtain the output values. The schematic of a fundamental building block of DNN is shown in Fig. 1 (a). It represents a neuron or a node in the hidden or output layers of the network presented in Fig. 1 (b). The data is processed layer by layer in a way that output from one layer is fed to a node of the next layer using a sum of weighted input data ($\Sigma w_i x_i$) and a non-linear activation function f. For simplicity, the bias terms are not shown here. The most common activation functions are linear ($f(x) = x$), sigmoid $\left(f(x) = \frac{1}{1 + e^{-x}}\right)$, rectified linear unit (ReLU) $\left(f(x) = \begin{cases} 0 \ for \ x \le 0 \\ x \ for \ x \ge 0 \end{cases}\right)$, and tanh $\left(f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}\right)$. The fully connected network shown in Fig. 1 (b) is called a feed-forward multilayer neural network, in which weights, biases, and activations functions are used to convert the inputs to the outputs. The DNNs can approximate any continuous function [19], making these suitable to use for demand forecasting when enough
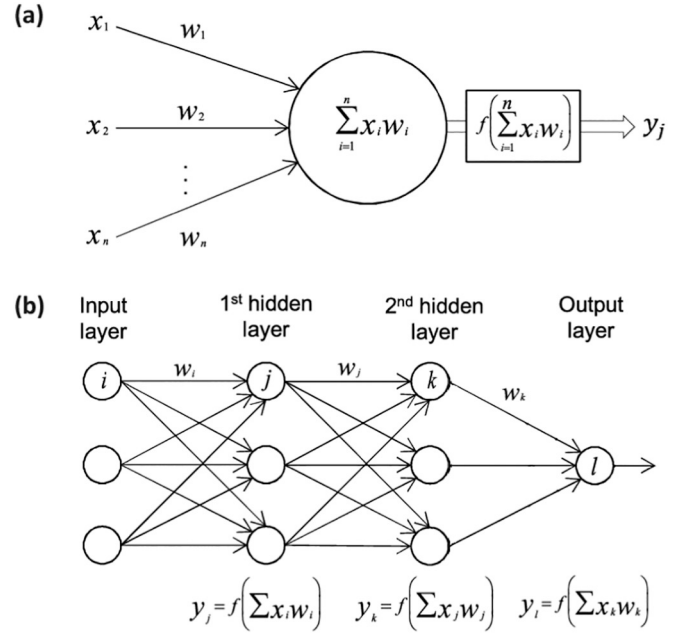

Fig. 1. Schematic diagram for a deep neural network – (a) the mathematical functions, and (b) A deep neural network architecture.

historical data is available. Recently DNNs are found to be useful for time-series forecasting problems [15,28]. Following this line of research, we develop ML/DL based demand estimation models to leverage the vast amount of data generated by a large number of products in a retail environment to generate accurate demand estimates.

3.3. Inventory optimization

The maximal approximation, empirical, and proposed data-driven optimization models are presented. The demand estimates from the previous section were used as the inputs and the optimal order quantity (q_i) is the expected outcome from these inventory optimization models.

3.3.1. The maximal approximation model
Scarf [35] provided the min-max model for inventory optimization when demand distributions are unknown, and only mean, and standard deviation of the demand is known to the user. Scarf's ordering rule provides the order quantity, assuming the worst and best possible cases of demand distribution with mean, μ, and standard deviation, σ. The detailed mathematical proof was provided (readers are suggested to refer to Scarf [35] for the details), which was later extended by the other authors [16]. The paper use this approximation model to calculate the order quantities and total cost. The results are used to benchmark the empirical and the proposed model.

3.3.2. Empirical demand distribution model
Based on the historical demand data, $S^i = \{(d_j^i, X_j^i)\}_{j=1}^n$, the mean demand forecasts are calculated. The forecast errors obtained from demand data and mean forecasts are obtained. The errors are assumed to follow a specific distribution, assumed normal distribution here. The parameters of the normal distribution – mean (μ_i) and standard deviation (σ_i), and thus, cumulative distribution function, $F^i(.)$, are calculated using forecast errors. The optimal solution to the inventory optimization problem, i.e., an optimal order quantity, Q_i, is obtained for a period, p, using the following equation:

$$Q_i^*(X_j) = \hat{d_i}(X_j) + inf\left\{y: F^i(p, \mu_i, \sigma_i) \ge \left(\frac{b^i}{b^i + h^i}\right)\right\} \tag{4}$$

Where $\hat{d_i}$ is the mean demand forecast, X_j are the independent features,

and $\frac{b^i}{b^i + h^i}$ is the optimal service level, which represents the probability of satisfying demand in a period, p.

3.3.3. Proposed inventory optimization model: QR-ML

The proposed approach uses the quantile regression - machine learning (QR-ML) methods and helps to reduce the burden of the calculating mean forecast, forecast errors, and assuming a specific distribution of the forecast errors. In this approach, the optimal order quantities are directly calculated using historical demand and independent features data for an item i, i.e. $\{d_j^i, X_j^i\}_{j=1}^n$. The proposed approach is based on the machine learning principle of empirical risk minimization (ERM) [38]. The ERM approach assumes that optimal order quantity is a linear function of the features and, thus, can be formulated as a linear program [5]. This ERM approach is extended to non-linear relationships by incorporating the ML methods, in the following way:

$$\min_{Q^i(\varphi^i X)} \hat{R}(Q^i(\varphi^i X); S^i) = \min_{q^i} \frac{1}{n} \sum_{j=1}^n [b^i(d_j^i - Q_j^i(\varphi^i X_j))^+ + h_j^i(Q_j^i(\varphi^i X_j) - d_j^i)^+], \quad (5)$$

where, $Q_j^i(\varphi^i, X_j)$ is the output in the period j with X_j as the input and φ^i are the parameters of the ML-based demand estimation method. The following problem can be reformulated using the dummy variables u_j^i and o_j^i for the underage and overage in period j. The non-linear program (NLP), newsvendor empirical risk minimization, is formulated as:

NV-ERM:

$$\min_{Q^i(\varphi^i X)} \hat{R}(Q^i(\varphi^i X); S^i) = \min_{q^i} \frac{1}{n} \sum_{j=1}^n [b^i(d_j^i - Q_j^i(\varphi^i X_j))^+ + h_j^i(Q_j^i(\varphi^i X_j) - d_j^i)^+]$$

$$\equiv \min \frac{1}{n} \sum_{j=1}^n (b^i u_j^i + h^i o_j^i)$$

subject to $\forall \; j = 1, 2, ..., n$

$$u_j^i \geq d_j^i - Q_j^i(\varphi^i, X_j)$$

$$o_j^i \geq Q_j^i(\varphi^i, X_j) - d_j^i, \qquad u_j^i, o_j^i \geq 0$$

The NV-ERM is an NLP of $p + 2n$ dimensional decision vector and $4n$ constraints. The objective of the NLP is to minimize the overall empirical risk ($\hat{R}$) of backordering and holding, i.e., the deviation of demand estimates from actual demand is correctly assigned to underage and overage. The objective function, $\hat{R}$ is a function of $Q^i(\varphi^i, X)$ with respect to the data, S^i. Using data, S^i, the parameters of the ML method (φ^{i*}) are obtained for the minimization of empirical risk, $\hat{R}$ in NV-ERM. The corresponding order quantity for item i in period j is the quantile forecast of $Q_j^i(\varphi^{i*}, X_j)$.

Ban and Rudin [5] and Huber et al. [22] showed that the objective function of the NV-ERM has the same form as that of loss function of high-dimensional quantile regression (QR) and thus, it can be solved using QR and the demand data. The loss function of QR is given by:

$$\mathcal{L}(\xi_j \mid \alpha) = \begin{cases} \alpha \xi_j & \text{if } \xi_j \geq 0 \\ (1 - \alpha)\xi_j & \text{if } \xi_j < 0 \end{cases} \quad (6)$$

where ξ_j is the error and $0 < \alpha < 1$. By using quantiles of underages ($\alpha \in (0,1)$) and overages $(1 - \alpha)$, an estimate for the quantile in inventory optimization can be obtained by using the following loss function, $\sum_{j=1}^n (\alpha(d_i - \widehat{d_i})^+ + (1 - \alpha)(\widehat{d_i} - d_i)^+)$ [25]. In this loss function, $\alpha^i = \frac{b^i}{b^i + h^i}$ of demand distribution or $(1 - \alpha^i) = \frac{h^i}{b^i + h^i}$ and using these values of α^i and $(1 - \alpha^i)$ in loss function, reproduces the NV-ERM. Thus, it now can be solved using quantile regression with the direct input of demand data and independent variables.
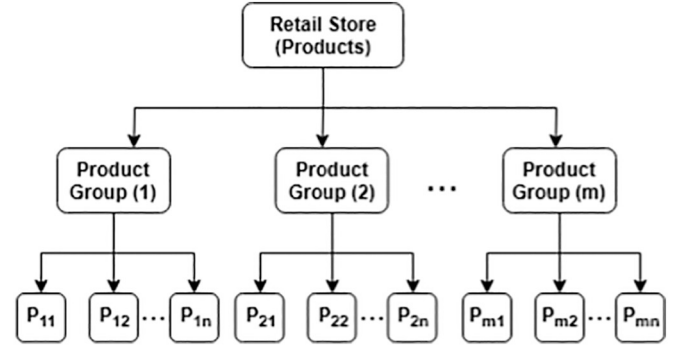

Fig. 2. Products' Hierarchy in retail store.

3.4. The multi-item inventory optimization with a capacity constraint

As mentioned in the literature, in retail stores, products are arranged into hierarchies where one product group breaks down into several of its variants (Fig. 2). This hierarchies are utilized to improve the accuracy and efficiency of the demand estimation algorithm in the literature [10,21,39] and in this study, it is used in solving the capacity constraint problem in multi-item inventory management in situations where a retailer has limited capacity (shelves) at the product group level, e.g., packaged foods of different flavors, etc.

In a product group (PG), as shown in Fig. 2, the products are of similar characteristics so that these can be assumed of a similar cost structure. By similar cost structure, it is meant that 1) underage cost of products is decided by a constant markup from the unit price, i.e. $b^i = m * c_i$, and 2) overage cost of products is a constant proportion of the unit cost, i.e. $h^i = c_i * d$. As mentioned in the literature, these two conditions are generally valid for products in a retail store [12,39]. So, the inventory optimization algorithm (QR-ML) is applied at product groups levels, and unconstrained optimal quantities are found out.

To derive the optimal order quantities (Q_i^*) at product level from unconstrained optimal quantities (Q_{PGi}^*) at the product groups, three possible cases can arise when a capacity constraint is active. Say, available capacity for each group is C_{PGi}. Then three possible cases are 1) $Q_{PGi}^* < C_{PGi}$, 2) $Q_{PGi}^* = C_{PGi}$, and 3) $Q_{PGi}^* > C_{PGi}$. For the first two cases, as the capacity constraint is already satisfied, and, it is proposed to use historical demand proportions as in top-down hierarchies to find the optimal quantities for products. The top-down is an efficient way of obtaining accurate forecasts [21]. The historical demand proportions are calculated from the available historical demand data [34].

For the third case, we used the mean and variance of items to find the optimal order quantities individually for all items. The works of Nahmias and Schmidt [29], Gallego and Moon, (1993), and Erlebacher [13] were followed to develop a data-driven heuristic solution. In general, multi-item NVP can be written as Z:

$$(Z) \min \; \pi = \sum_{k=1}^{G_n} \pi_k(Q_{kG})$$

subject to

$$\sum_{k=1}^{G_n} v_k(Q_k) \leq C, \text{ and } Q_k \geq 0 \; \forall \; k = 1, 2, ..., G_n.$$

With the assumption of the similar cost structure for the items under a PG, optimal service levels, $\alpha_k \left(= \frac{b^k}{b^k + h^k} \; \forall \; k = 1, 2, ..., G_n\right)$ are identical, and ratios of revenue per unit due to selling versus salvaging, $\beta_k \left(= \frac{v^k}{b^k + h^k} \; \forall \; k = 1, 2, ..., G_n\right)$, for all items will also be identical. So, we can write the following property regarding optimal order quantities of these k items.

Property 1. If $\alpha_k = \alpha$, and $\beta_k = \beta$, for $k = 1, 2, ..., G_n$, then $F_1(Q_1^*) = ... = F_k(Q_k^*) = ... = F_{G_n}(Q_{G_n}^*)$.

This property holds even when a capacity constraint is present and for any set of demand distributions of items. Based on *Property 1*, we can get optimal order quantities for all items from the following theorem.

Theorem 1. Suppose $l_1, l_2, ..., l_m$ are the common shape parameters of the G_n demand distributions and transformation T is defined as $T(d_i, \mu_i, \sigma_i, l_1, l_2, ..., l_m, f, g) = \frac{d_i - [\mu_i - \sigma_i f(l_1, l_2, ..., l_m)]}{\sigma_i g(l_1, l_2, ..., l_m)}$, where f, g are the functions of the shape parameters. Under the capacity constraint and condition that $T(d_i, \mu_i, \sigma_i, l_1, l_2, ..., l_m, f, g) \sim T(d_j, \mu_j, \sigma_j, l_1, l_2, ..., l_m, f, g) \forall (i, j) \in \{1, 2, ..., G_n\}$, the optimal quantity is given by

$$Q_k^* = \mu_k + \sigma_k \frac{C - \sum_{k=1}^{G_n} v_j \mu_j}{\sum_{k=1}^{G_n} v_j \sigma_j} \tag{7}$$

Proof: Appendix I.

Theorem 1, under the assumption of products *i, j* have identical distribution (denoted by ∼ in Theorem 1), provides the conditions to get the optimal order quantities for the products of a PG. Note that the expression for optimal quantities is free from any specific distribution assumption; hence, we accomplish our goal of providing a data-driven solution to the multi-item inventory optimization.

Overall, a data-driven solution to multi-item inventory optimization is provided using the proposed QR-ML approach and a heuristic. In the proposed framework, the computation efforts will get reduced by a factor of *N/M* in an ideal case when all forecasts of PGs are less than capacity, where N is the number of products, M is the number of product groups. This novel approach is used to solve a multi-item NVM using top-down hierarchies; thus, we may call it a top-down NVM (TD-NVM). The performance of the proposed approach is investigated in detailed empirical analysis in the next section.

4. Empirical evaluation

In this section, first, the data used for empirical analysis are described in detail. After that, the basic details and parameter setup for time-series methods, i.e., seasonal naïve, ETS, ARIMA, and ARIMA with external regressors; and machine learning methods, i.e., multiple regression, neural networks, and random forest are discussed. The different forecasting methods are evaluated on multiple performance criteria to judge the best demand forecasting method. The impact of these methods on inventory cost is analyzed. After that, inventory optimization techniques - maximal approximation, the empirical optimization (normal distribution), and the proposed method – quantile regression - machine learning (QR-ML) method, as explained in Section 3.3, are applied. The results from these inventory optimizations are analyzed to select the best model based on minimum inventory cost.

4.1. Data and descriptive statistics

The real-world data is collected from a retailer who sells perishable food items. The dataset consists of point-of-sales data for 16 such items and is available for 156 weeks. The retailer has to decide the order quantities for each item per week, and in this way, the business requires at least 16 inventory decision makings every week. The time-series data consist of multiple independent variables related to point-of-sales, promotions, store, calendar, holidays, weather, etc. The summary statistics of the variables from the dataset are provided in Table 2.

Here economic activity index (Econ_Index) measures the average economic growth in a metropolitan area [2] and found to be a good indicator of sales in the literature [4]. The data for weather and economic indicators is collected through the website of the National Oceanic and Atmospheric Administration (NOAA), USA. Four weeks of data were used as the test dataset, and the rest of the data is used as the training dataset.

Table 2
Summary statistics of variables.

	Descriptive statistics		
Variable name	Median	Mean (SD)	Remark
Sales	40	42.89 (14.61)	Product sales
Price	2.89	3.20 (1.50)	Selling price
% Discount	0.00	0.22 (0.416)	Promotional discount
Visits	17	20.09 (12.64)	No. of customer visits
HHS	17	19.72 (12.33)	No. of households
TAVG	497	490.8 (89.39)	Average temperature
TMIN	429.5	424.9 (93.73)	Minimum temperature
TMAX	563	556. 7 (87.30)	Maximum temperature
PRCP	0.1850	0.756 (1.33)	Precipitation
Eco_Index	3.41	1.183 (5.40)	Economic Index
Oil_Price	2.754	2.896 (0.53)	Oil Price
	% of cases		Remark
Feature	95.35		In in-store circular
Display	7.57		In-store Display
TPR	9.81		Temporary price reduction

4.2. Forecasting methods, parameter setup, and performance metrics

The widely used time-series methods for sales forecasting like seasonal naïve, ETS, and ARIMA (Auto-Regressive Integrated Moving Average); and machine learning methods like multiple regression, artificial neural network, and random forest, are used as the benchmarking forecasting methods. Also, we provide the details and parameter selection of these methods for our forecasting purposes.

4.2.1. Time-series methods

In the seasonal naïve method, actual demand from some last period is used as the forecast for the next period. The naïve method is used for forecasting purposes where the next period's demand is strongly dependent on some previous period's demand. The forecast for the next period is equal to actual values from the kth period of the past season, i.e. $\hat{y}_{t+1} = y_{t+1-k}$. In our case, we used the frequency of $k = 4$. The number was selected based on the acf plots of the residuals from the naïve forecasts as k = 4 resulted in non-significant autocorrelation in the acf plots of residuals.

ETS is a class of exponential smoothing methods that are used to model the error, trend, and seasonality. A simple ETS method, single exponential smoothing, predicts the next period's demand based on a weighted sum of previous observations. The advanced versions, double and triple exponential smoothing, use the trend and seasonality also to produce more accurate forecasting results. The *forecast* package [23] is used to estimate the parameters of the models for demand forecasting in our case.

ARIMA (Auto-Regressive Integrated Moving Average) method is widely applied for demand forecasting in the supply chain [17]. ARIMA model combines the autoregressive (AR) and moving average (MA) model of time series forecasting. AR(p) model uses past observations to predict future values. The AR model is defined as: $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + e_t$. y_t is regressed over *p* time-lagged value of y_t, and e_t, which is the white noise in the time series. ϕ_1 and *c* are the model parameters. Similarly, in MA(q) model is defined as: $y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} + e_t$. In MA(q) model, past forecast error values are used rather than forecast variable values. y_t is regressed over *q* time-lagged errors and, θ_1 and c are model parameters. Overall, ARIMA (p, d, q) is defined as: $y'_t = c + \phi_1 y'_{t-1} + ... + \phi_p y_{t-p}' + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} + e_t$. The y'_t represents the differenced value of the dependent variable. p, q, and *d* are the parameters of the ARIMA model. d is the degree of differencing. ARIMA model is fitted, and optimal parameters of the ARIMA model are estimated such that errors are minimized. The Box-Jenkins method with Akaike Information Criterion (AIC) and Bayesian Information Criterion

Table 3
Error metrics for demand estimation.

	ME	MAE	RMSE	MAPE	Rel-ME	Rel-MAE	Rel-RMSE
s-naive	−1.8854	7.4271	8.9059	50.5903	1.0000	1.0000	1.0000
ARIMA	−1.5737	5.5223	6.7977	42.6875	0.8347	0.7435	0.7633
ETS	−3.0334	6.6520	7.9669	51.3972	1.6089	0.8956	0.8946
FNN	−0.4626	5.4530	7.1420	44.1978	0.2454	0.7342	0.8019
RF	0.0281	1.4695	1.9264	8.7740	−0.0149	0.1979	0.2163
DNN	−0.0356	1.7121	2.2545	12.1680	0.0189	0.2305	0.2531
ARIMAx	0.3245	2.5112	3.3312	17.3360	−0.1721	0.3381	0.3740

Table 4
FVA Analysis.

	Forecast accuracy (%)	Comparative improvements in forecast accuracy (%)	Cumulative FVA over benchmark, s-naïve (%)
s-naive	49.40	–	–
FNN	55.80	6.40	6.40
ARIMA	57.31	1.51	7.91
ARIMAx	82.66	25.35	33.26
DNN	87.83	5.17	38.43
RF	91.22	3.39	41.82

Table 5
Average inventory costs relative to the best approach for each product groups.

Demand Forecast Method		PG1	PG2	PG3	PG4
		Cost	Cost	Cost	Cost
s-naive	MA	2.78	1.89	4.28	9.24
	Norm	2.22	1.51	3.42	7.37
ARIMA	MA	3.45	1.44	5.68	2.07
	Norm	2.75	1.34	4.53	2.01
ARIMAx	MA	1.63	1.25	5.84	2.58
	Norm	1.30	**1.00**	4.66	2.06
ETS	MA	2.69	1.78	9.96	4.12
	Norm	2.14	1.42	7.93	3.29
FNN	MA	2.78	1.89	6.46	6.74
	Norm	2.22	1.51	4.28	5.38
	QR	1.81	1.74	4.16	2.54
RF	MA	1.30	2.50	2.02	1.28
	Norm	1.11	2.33	1.74	1.03
	QR	**1.00**	2.32	**1.00**	**1.00**
DNN	MA	1.59	4.38	3.42	1.69
	Norm	1.27	3.19	1.75	1.65
	QR	1.46	3.77	1.39	1.54

Bold letters (1.00) indicate the method with minimum inventory cost and other costs in the same column are calcualted relative to it.

(BIC) as error minimization criteria were used to find the values of parameters *p, d, and q*.

4.2.2. Machine learning methods

4.2.2.1. Feedforward Neural Networks (FNNs).
Single-layer feed-forward neural networks are widely used to model time-series forecasting [40]. The FNN has three types of layers, namely, an input layer, an output layer, and hidden layers. The number of hidden layers was restricted (equal to 1) as with more than one hidden layer configurations are covered by DNN. The inputs $(X_1, X_2, ..., X_p)$ and outputs (y_t) are related through the equation: $y_t = \alpha_o + \sum_{j=1}^{q} \alpha_j g (\beta_{oj} + \sum_{i=1}^{p} \beta_{ij} X_{t-i}) + \varepsilon_t$, where αs and βs are connection weights, p is the number of input nodes and q is the number of hidden nodes. The final output of the network is $y_t = f (X_1, X_2, ..., X_p, w) + \varepsilon_t$, where w is a vector of all parameters, and f is a non-linear function that maps the inputs to outputs with the help of w and connection weights. Thus, the neural network is equivalent to a nonlinear model. In our model, while architecting the neural network, time-series data was de-trended and deseasonalized through differencing, and then we used the single-layer feed-forward neural network.

4.2.2.2. Random Forest (RF).
The details of the RF algorithm are already explained in Section 3.2, so only important parameters of RF are discussed here. There are two main characteristics of the RF, namely, the generalization error and the measure of variable importance (MOVI). The former is also known as out-of-bag-error (OOBE) and is the estimation of the generalization ability of the model. It is the mean prediction error of the out of sample or first-seen observations and serves the purpose of internal cross-validation. The OOBE is given by $OOBE = \frac{1}{n} \sum_{i=n}^{n} (Y_i - \widehat{Y_i})^2$. On the other hand, the MOVI is obtained by permuting a feature and averaging the difference in OOBE before and after permuting over all trees, $MOVI(X^j) = \frac{1}{q} \sum_{l=1}^{q} (\widehat{OOBE_l} - OBBE_l)$. The higher the value of *MOVI* (X^j), the more critical is feature X^j. Through OOBE and MOVI, the RF provides the best fit among independent variables and targets. The parameters are explained in Section 4.3.

4.2.3. Performance metrics
The performance of the forecasting methods is compared using three essential characteristics. These are:

- Bias – to check the forecasting method for its tendency of over-forecasting or under-forecasting of actual values.
- Accuracy – to check how closely forecasted values conforms to the actual values.
- Uncertainty – to check the average deviation of the forecast from the mean forecast.

The performance metrics used to measure the characteristics mentioned above are as follows: mean error (ME), mean absolute error (MAE) and root mean squared error (RMSE) for bias, accuracy, and variance respectively [24]. The practice-oriented accuracy metric, mean absolute percentage error (MAPE), is also included in the analysis. The multiple performance metrics are used because each of these metrics has some strengths and weaknesses. For example, MAE and RMSE are scale-dependent. MAPE is bounded on a lower limit and unbounded on the upper limit.

As the errors are calculated for multiple demand series, these errors are aggregated to get a single metric to compare the forecasts known as average error. Because the performance of different forecasting methods is compared, therefore, the relative measures are also included. The relative metrics are easy to interpret, scale, and allows to summarize the results from different series concisely and coherently [11,36]. The relative mean error, relative mean absolute error, and relative mean squared errors are used to compare the performance of these methods. These relative errors are calculated by dividing the sum or mean of errors from the evaluated method by that of a benchmark method. The seasonal naïve method is used as a benchmark method.

Also, a practice-oriented approach, FVA (forecast value-added) analysis is performed to compare the forecasting methods based on MAPE [18]. FVA helps to identify percentage improvement by a method
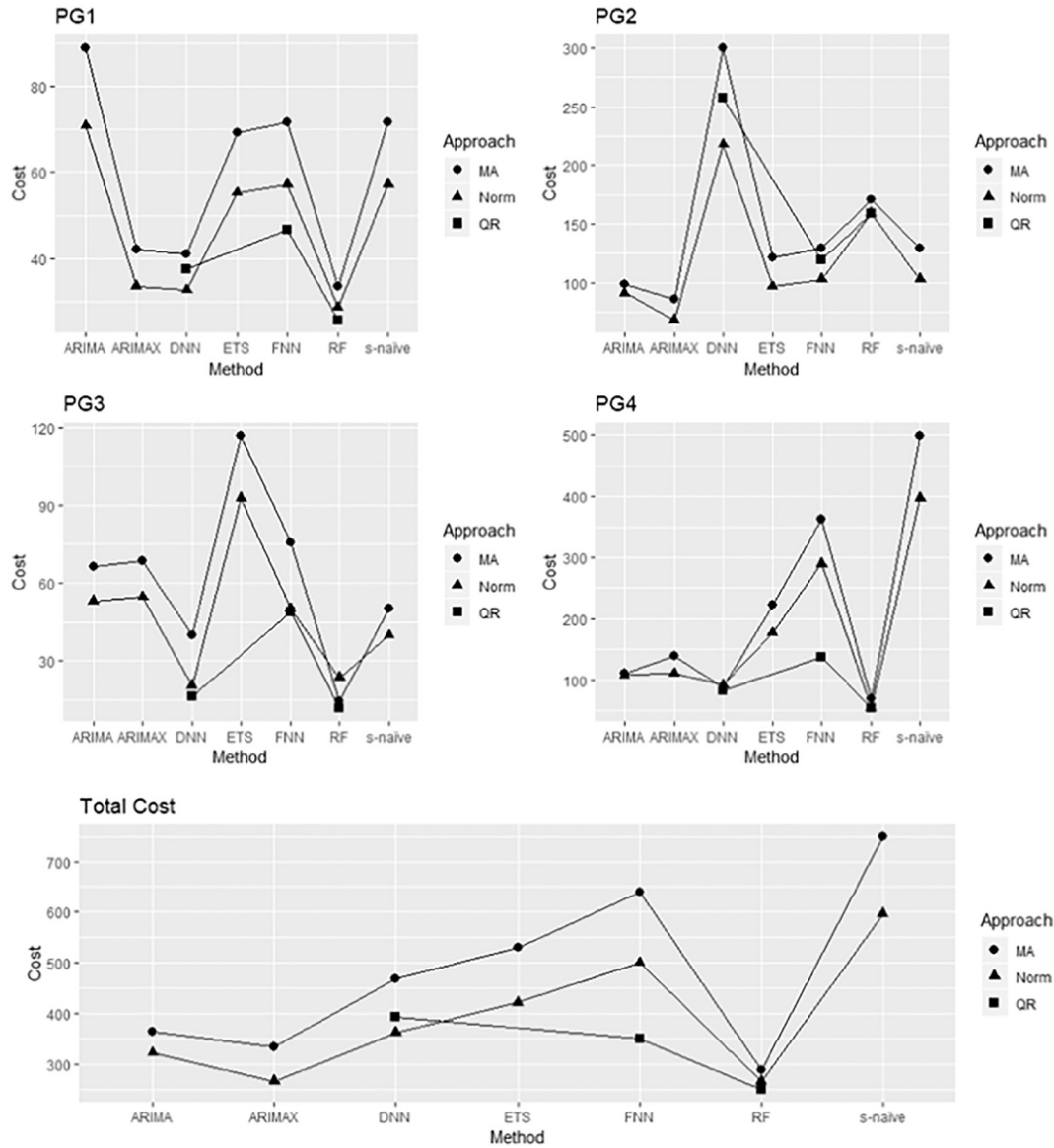
Fig. 3. Comparison of inventory cost from different inventory optimization techniques using different demand estimation techniques.

over other methods and benchmark methods. FVA consists of following three steps approach:

1) Calculate forecasting accuracy using MAPE, i.e. (100 – MAPE),
2) Estimation of FVA points to know the performance of the proposed model over the benchmarking model using the following equation:
   FVA = (forecast accuracy of proposed method) − (forecast accuracy of benchmark method)
3) Estimation of marginal improvements in forecasting accuracy by a different method.

In this way, the following metrics are used for the analysis: mean error, mean absolute error, root mean squared error, mean absolute percentage error, the relative average mean error, the relative average mean absolute error, relative average root mean squared error and FVA analysis.

4.3. Demand estimation: Results and discussions

To implement the forecasting models, parameter tuning was

performed. The k = 4 for the naïve method is selected based on acf plots of the residuals. The auto.arima and ets functions were used for auto-selection of the parameters, and the seasadj function is used for the deseasonalized series for FNN from the forecast package in R language [23]. For the optimization of the number of nodes, the grid search algorithm is used in FNN. For RF, repeated cross-validation with a random search grid is used to find the best set of parameters [15]. The maximum depth = 20, and the number of trees were allowed to be 500, and the number of variables randomly sampled as candidates at each split (mtry) was left default as $\sqrt{p}$, where p is the number of variables [8]. For DNNs, we used the network with three hidden layers with a different set of nodes varying from 12 to 36 (and input layer of 12 and output layer of 1 node) were used and optimal configuration was selected by hyperparameter optimization performed by designing a grid search within Keras [6]. Besides, dropout of {01–0.5} and L1 regularization was used to avoid overfitting.

The results from the demand estimations are presented in Table 3. Three months of data is used as test data for the calculation of out-of-sample-errors. The bias (ME), accuracy (MAE) and variance (RMSE), and other measures of accuracy of the forecasts are presented. It is
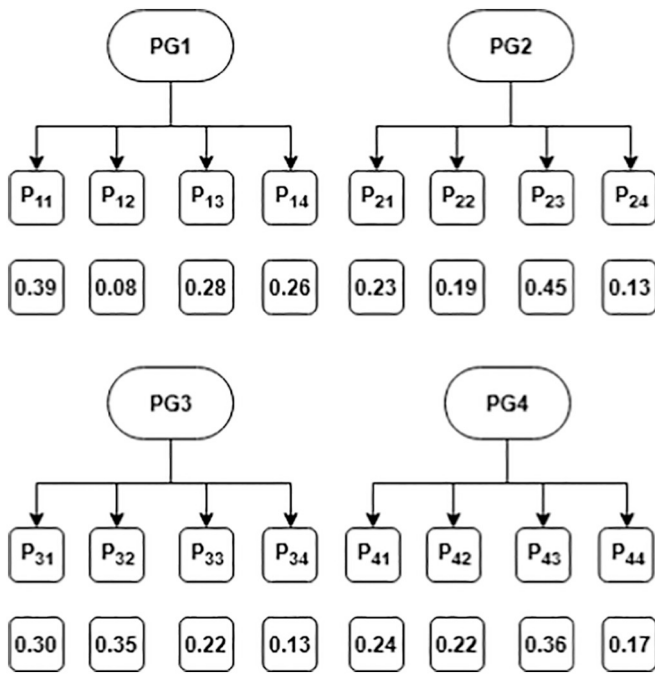
Fig. 4. Products hierarchies and their historical demand proportions.

observed that mean errors (bias) for the methods (except RF and ARIMAx) are negative, which means that these methods are underestimating the demand. RF has the smallest mean error, i.e. on average, predictions from RF are neither underestimated nor overestimated, and thus, it can be said that RF has a minimum bias in prediction. MAEs suggests that the accuracy of machine learning methods is higher than time-series methods. Notably, RF and DNN are significantly more accurate than other methods. A similar trend is exhibited in RMSE, which means RF, and DNN have the least variance, too, among all methods.

During analysis, it was observed that forecasting of sales pattern of a few products is dependent on the time-based components and as well as on independent variables. The hybrid methods are suitable for this type of pattern prediction. Thus, for the proof of the concept, a state-of-the-art hybrid method, ARIMAx, is included in the analysis. ARIMAx is an extended version of the ARIMA which considers the independent variables too. ARIMAx is a hybrid of ARIMA and multiple linear regression, and its advantage is that it uses the autocorrelation present in the residuals, and regression between demand and independent variables to improve the forecasting accuracy of the model. As evident from Table 3, ARIMAx is performing better than the time-series method due to the added prediction power of regression. Punia et al. [31] can be referred for a detailed study on hybrid machine/deep learning methods.

Overall, RF and DNN are better than other methods on all three characteristics viz. bias, accuracy, and variance. MAPE results also established that RF is best among all methods. For a concise and scale-independent comparison of the error metrics, the relative errors viz. Rel-ME, Rel-MAE, and Rel-RMSE are also presented in Table 3.

The value-added analysis for forecasts in Table 4 suggests that except ETS, all other methods have better forecast accuracy than the s-naïve, the benchmark method. The comparative improvements in FVA analysis show a significant jump in forecast accuracy from ARIMA (a

time-series method) to ARIMAx, which highlights the significance of using external business information in the data-driven prediction models. DNN and RF are the best and have small but significant incremental improvements in the forecast accuracy over ARIMAx.

4.4. Inventory analysis

For inventory analysis, we are using data for four product groups consisting of a total of 16 products. As explained in Section 3.4, these products can be mapped to the structure of a products' hierarchy. The products in each group exhibit very similar characteristics. The inventory analysis is initiated with each group, which later taken to the product level with the help of top-down product hierarchies and proposed heuristic.

4.4.1. Effects of demand estimation on inventory costs

The demand estimates for the next period in the newsvendor model for each product group (PG) are used in the NVM to calculate the inventory costs. The inventory costs are calculated at the optimal service level $\left(\frac{b^i}{b^i + h^i}\right)$. The average relative inventory costs for different product groups are presented in Table 5.

The machine learning methods (RF and DNN) have better performance than time-series methods in all cases except PG2, where ARIMAx is best. The better performance of ML methods can be attributed to the large data used by these methods to produce more robust and accurate forecasts. These robust forecasts lead to better inventory costs. The results from the PG2 are showing an aberration, which further supports our proposition in Section 4.4 that some demand patterns of some products can be better modeled through a hybrid method. As here, ARIMAx has a lower cost than time-series and ML methods. Overall, from Table 5, it can be concluded that demand estimation significantly impacts inventory costs.

4.4.2. Effects of inventory optimization techniques on inventory cost

The inventory costs from the empirical method with normal distribution and proposed QR-ML method are compared with the cost from the maximal approximation approach. The average inventory costs for each product group are calculated and plotted in Fig. 3.

The plots in Fig. 3 are analyzed in two different dimensions. First, along with demand estimation techniques, the effect of accurate demand estimations is dominating the marginal impact from inventory optimization techniques. Second, it can be observed, among inventory optimization models, that normal distribution based NVM has a lower cost than MA in all of the cases, and proposed QR-ML based inventory optimization has a lower cost than normal distribution based empirical model except in a few cases where the difference is relatively small and non-significant. Therefore, the proposed approach of quantile regression-based machine learning inventory optimization is performing better than other combinations. It is because of the prediction power of the ML method and efficient use of data for optimization in-place of assuming erroneous demand distributions for inventory optimization.

4.4.3. The multi-item capacity allocation

After finding the minimum cost unconstrained quantities for PGs, the product level optimal quantities are derived in this section. The available capacity for each product group is 200 units in the store. The optimal order quantity for each of the products is estimated through

Table 6
Optimal Order quantities.

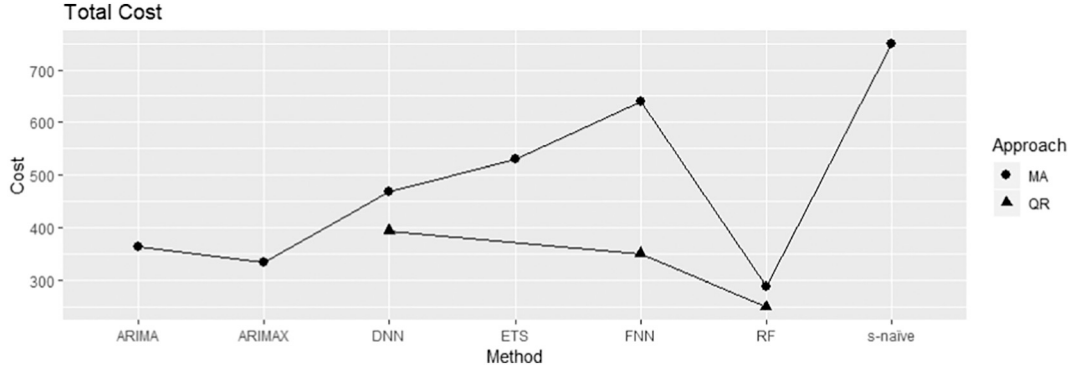| Product Group	PG1				PG2				PG3				PG4			
Products	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Order Quantities	75	15	54	50	46	38	90	26	60	64	41	35	45	42	67	32

Fig. 5. Comparison of total inventory cost from the inventory optimization techniques and demand estimation techniques.

QR-RF based inventory optimization technique. The order quantities found out to be 192, 199, 213, and 184 units for PG1, PG2, PG3, and PG4, respectively. Overall, the total quantity is 788, which is under the total available capacity of 800 units. However, at the PG level, only three optimal quantities (PG1, PG2, and PG4) are under the available capacity for an individual PG.

To apply top-down allocation of capacity to products, the historical proportion is calculated for all product hierarchies. The hierarchies and historical demand proportion are shown in Fig. 4. Except for PG3, direct order quantities can be found out by direct allocation, i.e. $Q^{i*} = p * Q_G^*$, where p is the proportions and Q^{i*} is the optimal order quantities for a product of a product group G, and Q_G^* is the optimal aggregate quantity for a product group G. The order quantities for products of PG1, PG2, and PG4 are presented in Table 6 (rounded off to higher numbers). For PG3, using the mean and standard deviation of the items in Eq. (7), the optimal quantities are found out and are also presented in Table 6.

Finally, the total cost is calculated for all products and plotted for three inventory optimization techniques within the multi-item case. As shown in Fig. 5, the total cost for the proposed approach QR-RF is found to be lower than the multi-item approach of Gallego and Moon [16]. Also, the effect of accurate demand estimations is clearly dominating the effect of inventory optimization techniques.

Overall, through the above analyses, it can be said that: 1) the proposed machine learning methods are significantly better than other benchmarking methods for demand estimation, and 2) the performance of proposed QR-ML inventory optimization techniques found to be better than other inventory optimization techniques approaches. Also, the proposed heuristic provided the optimal order quantities for the products within capacity constraint from already minimum cost order quantities of their respective product groups.

5. Conclusion

In this paper, we proposed a data-driven solution to a multi-item newsvendor model with a capacity constraint. The proposed model uses the machine learning-based quantile regression to obtain the order quantities for the newsvendor model in a single step. Thus, it eliminates the need for separate demand estimation and inventory optimization processes. Besides, it includes a heuristic to provide the order quantities under competitive resource requirements among multiple products. Through the proposed method, we filled two crucial research gaps in the literature on data-driven multi-item inventory optimization. The first gap was to use advanced non-linear machine learning methods with quantile regression to find unconstrained order quantities. The second gap was to extend the data-driven model to solve the multi-item newsvendor problems with a capacity constraint. Filling these gaps will help in solving the demand forecasting and inventory management problems, which are relevant to many industries, such as fast-food, fashion clothing, groceries, bakeries, sports goods, and perishable products industry.

The performance of the proposed method is investigated through several analyses, and the following important observations are worth mentioning here. These are: 1) the external variables play a significant part in enhancing the accuracy of the demand estimation method, 2) the machine and deep learning methods outperformed the time-series methods for demand estimation with large datasets, 3) the accurate demand estimations have significant impact on total cost of the inventory management, and 4) the proposed data-driven QR-ML approach is better than the empirical solution.

The data-driven models incorporating machine learning methods for the newsvendor model are an active field of research; thus, the research can be extended in several directions. As come up during analyses, the use of hybrid demand estimation methods and their incorporation in quantile regression's loss function will be a good opportunity to explore. Especially, it includes designing more advanced deep learning-based method to develop the data-driven model for newsvendor problem. Finally, solving multi-period newsvendor models using data-driven methods will be an exciting area for future research.

Appendix A. Appendix I

Since $T(d_i, \mu_i, \sigma_i, l_1, l_2, ..., l_m, f, g) \sim T(d_j, \mu_j, \sigma_j, l_1, l_2, ..., l_m, f, g) \ \forall \ (i,j) \in \{1, 2, ..., G_n\}$, and

$F_i(Q_i^*) = F_i(Q_j^*) \ \forall \ (i,j) \in \{1, 2,..., G_n\}$

therefore, $\dfrac{Q_i^* - [\mu_i - \sigma_i f(l_1, l_2, ..., l_m)]}{\sigma_i \, g(l_1, l_2, ..., l_m)} = \dfrac{Q_j^* - [\mu_j - \sigma_j f(l_1, l_2, ..., l_m)]}{\sigma_j \, g(l_1, l_2, ..., l_m)}, \ or$

$Q_j^* = \mu_j - \sigma_j f(l_1, l_2, ..., l_m) + \dfrac{\sigma_j}{\sigma_i} * (Q_i^* - [\mu_i - \sigma_i f(l_1, l_2, ..., l_m)])$

Multiply both sides by v_j and aggregate over j,

$\sum_{j=1}^{G_n} v_j Q_j^* = \sum_{j=1}^{G_n} v_j \left[\mu_j - \sigma_j f(l_1, l_2, ..., l_m) + \dfrac{\sigma_j}{\sigma_i} * (Q_i^* - [\mu_i - \sigma_i f(l_1, l_2, ..., l_m)]) \right]$

Given, $\sum_{j=1}^{G_n} v_j Q_j^* = C$, where C is the capacity for items under a PG, hence,

$$Q_i^* = \mu_i + \frac{\sigma_i}{\sum_{j=1}^{G_n} v_j \sigma_j}\left(C - \sum_{j=1}^{G_n} v_j \mu_j\right).$$

Or, for a Kth product of a PG, $Q_k^* = \mu_k + \sigma_k \dfrac{C - \sum_{k=1}^{G_n} v_j \mu_j}{\sum_{k=1}^{G_n} v_j \sigma_j}$ QED.

References

[1] I. Alon, M. Qi, R.J. Sadowski, Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods, J. Retail. Consum. Serv. 8 (3) (2001) 147–156.

[2] M.A. Arias, C.S. Gascon, D.E. Rapach, Metro business cycles, J. Urban Econ. 94 (2016) 90–108, https://doi.org/10.1016/j.jue.2016.05.005.

[3] K.-F. Au, T.-M. Choi, Y. Yu, Fashion retail forecasting by evolutionary neural networks, Int. J. Prod. Econ. 114 (2) (2008) 615–630, https://doi.org/10.1016/j.ijpe.2007.06.013.

[4] F. Badorf, K. Hoberg, The impact of daily weather on retail sales: an empirical study in brick-and-mortar stores, J. Retail. Consum. Serv. 52 (2020) 101921.

[5] G.-Y. Ban, C. Rudin, The big data newsvendor: practical insights from machine learning, Oper. Res. 67 (1) (2018) 90–108, https://doi.org/10.1287/opre.2018.1757.

[6] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (2012) 281–305 (Retrieved from Scopus).

[7] D. Bertsimas, A. Thiele, A robust optimization approach to inventory theory, Oper. Res. 54 (1) (2006) 150–168.

[8] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[9] X. Chen, M. Sim, P. Sun, A robust optimization perspective on stochastic programming, Oper. Res. 55 (6) (2007) 1058–1071, https://doi.org/10.1287/opre.1070.0441.

[10] B.J. Dangerfield, J.S. Morris, Top-down or bottom-up: aggregate versus disaggregate extrapolations, Int. J. Forecast. 8 (2) (1992) 233–241.

[11] A. Davydenko, R. Fildes, Measuring forecasting accuracy: the case of judgmental adjustments to SKU-level demand forecasts, Int. J. Forecast. 29 (3) (2013) 510–522.

[12] Q. Ding, C. Dong, Z. Pan, A hierarchical pricing decision process on a dual-channel problem with one manufacturer and one retailer, Int. J. Prod. Econ. 175 (2016) 197–212, https://doi.org/10.1016/j.ijpe.2016.02.014.

[13] S.J. Erlebacher, Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint, Prod. Oper. Manag. 9 (3) (2000) 303–318.

[14] K.J. Ferreira, B.H.A. Lee, D. Simchi-Levi, Analytics for an online retailer: demand forecasting and Price optimization, Manuf. Serv. Oper. Manag. 18 (1) (2016) 69–88, https://doi.org/10.1287/msom.2015.0561.

[15] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, Eur. J. Oper. Res. 270 (2) (2018) 654–669, https://doi.org/10.1016/j.ejor.2017.11.054.

[16] G. Gallego, I. Moon, The distribution free newsboy problem: review and extensions, J. Oper. Res. Soc. 44 (8) (1993) 825–834, https://doi.org/10.1057/jors.1993.141.

[17] V. Gaur, A. Giloni, S. Seshadri, Information sharing in a supply chain under ARMA demand, Manag. Sci. 51 (6) (2005) 961–969.

[18] M. Gilliland, U. Sglavo, Worst practices in business forecasting, Analytics (2010) 12–17.

[19] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249–256.

[20] G.A. Godfrey, W.B. Powell, An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution, Manag. Sci. 47 (8) (2001) 1101–1112, https://doi.org/10.1287/mnsc.47.8.1101.10231.

[21] C.W. Gross, J.E. Sohl, Disaggregation methods to expedite product line forecasting, J. Forecast. 9 (3) (1990) 233–254, https://doi.org/10.1002/for.3980090304.

[22] J. Huber, S. Müller, M. Fleischmann, H. Stuckenschmidt, A data-driven newsvendor problem: from data to decision, Eur. J. Oper. Res. 278 (3) (2019) 904–915, https://doi.org/10.1016/j.ejor.2019.04.043.

[23] R.J. Hyndman, Y. Khandakar, Automatic time series forecasting: the forecast package for R, J. Stat. Softw. 27 (3) (2008) 1–22, https://doi.org/10.18637/jss.v027.i03.

[24] R.J. Hyndman, A.B. Koehler, Another look at measures of forecast accuracy, Int. J. Forecast. 22 (4) (2006) 679–688.

[25] R. Koenker, Quantile regression, (2005), https://doi.org/10.1017/CBO9780511754098.

[26] R. Levi, G. Perakis, J. Uichanco, The data-driven newsvendor problem: new bounds and insights, Oper. Res. 63 (6) (2015) 1294–1306, https://doi.org/10.1287/opre.2015.1422.

[27] R. Levi, R.O. Roundy, D.B. Shmoys, Provably near-optimal sampling-based policies for stochastic inventory control models, Math. Oper. Res. 32 (4) (2007) 821–839, https://doi.org/10.1287/moor.1070.0272.

[28] A.L.D. Loureiro, V.L. Miguéis, L.F.M. da Silva, Exploring the use of deep neural networks for sales forecasting in fashion retail, Decis. Support. Syst. 114 (2018) 81–93, https://doi.org/10.1016/j.dss.2018.08.010.

[29] S. Nahmias, C.P. Schmidt, Efficient heuristic for the multi-item newsboy problem with a single constraint, Naval Research Logistics Quarterly 31 (3) (1984) 463–474, https://doi.org/10.1002/nav.3800310311.

[30] A. Oroojlooy, L. Snyder, M. Takáč, Applying deep learning to the newsvendor problem, IISE Transac. (2019) 1–39 just-accepted.

[31] S. Punia, K. Nikolopoulos, S. Singh, J. Madaan, K. Litsiou, Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail, Int. J. Prod. Res. (2020) 10.1080/00207543.2020.1735666.

[32] S. Punia, S. Singh, J. Madaan, Deep learning and cross-temporal hierarchies based framework for demand forecasting, Proceedings of XXIII Annual International Conference of the SOM 2019, 2019, p. 62 Kanpur, India.

[33] Y. Qin, R. Wang, A.J. Vakharia, Y. Chen, M.M. Seref, The newsvendor problem: review and directions for future research, Eur. J. Oper. Res. 213 (2) (2011) 361–374.

[34] G. Sbrana, A. Silvestrini, Forecasting aggregate demand: analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework, Int. J. Prod. Econ. 146 (1) (2013) 185–198.

[35] H. Scarf, A min-max solution of an inventory problem, Studies in the Mathematical Theory of Inventory and Production, 1958, pp. 201–209.

[36] S. Shankar, P.V. Ilavarasan, S. Punia, S.P. Singh, Forecasting container throughput with long-short-term-memory networks, Ind. Manag. Data Syst. (2019), https://doi.org/10.1108/IMDS-07-2019-0370.

[37] E.A. Silver, D.F. Pyke, D.J. Thomas, Inventory and Production Management in Supply Chains, CRC Press, 2016.

[38] V.N. Vapnik, Statistical Learning Theory, (1998).

[39] M.A. Villegas, D.J. Pedregal, Supply chain decision support systems based on a novel hierarchical forecasting approach, Decis. Support. Syst. 114 (2018) 29–36, https://doi.org/10.1016/j.dss.2018.08.003.

[40] G.P. Zhang, An investigation of neural networks for linear time-series forecasting, Comput. Oper. Res. 28 (12) (2001) 1183–1202.

Sushil Punia (M.Tech (IIT Kanpur), B.Tech (NIT KKR)) is a Ph.D. candidate at the Indian Institute of Technology Delhi, India. His area of research in Ph.D. is the operations and supply chain management and he is working on machine learning and deep learning-based data-driven decision models for supply chains. His research has appeared in International Journal of Production Research (IJPR), Industrial Management and Data Systems (IMDS), etc. and presented at international conferences like INFORMS 2017, and SOM 2019. He has been awarded one of the Springer Nature's Best Paper Award at SOM 2019. He also awarded Emerald Outstanding Reviewer Award 2018. His research interests are application of AI in Healthcare, and Industry 4.0.

Surya Prakash Singh is Professor in the Department of Management Studies, IIT Delhi, India. He holds a PhD Degree from IIT Kanpur. He is also Post-Doctoral Fellow from NUS Singapore-MIT USA Alliance. His research interest lies in facility layout problems, heuristics and meta-heuristics. His work has been published in leading international journals such as IJPR, LNCS, IJAMT, EJM, RBR, IJRTE and APMR. He regularly reviews articles for many leading journals. His biography appeared in Marquis, USA Who's Who in Science and Engineering, December 2007, and Who's Who in the World, November 2010. Recently, he has been awarded Young Outstanding Faculty Fellowship from IIT Delhi.

Dr. Jitendra Madaan received his B.Tech Degree in Production and Industrial Engineering from M.B.M Govt. Engg College, Jodhpur (JNV University), India, and obtained his M. Tech in Manufacturing System Engg. from Department of Mechanical Engg. MREC (Now MNIT), Jaipur and PhD in Mechanical Engineering from the Indian Institute of Technology (IIT) Delhi, India. His current research interests are Reverse Logistics and Supply Chain Management, Sustainable Operations Management, Production Management, information and governance effectiveness, Systems Modelling and Simulation, etc. To date, Dr. Madaan has published over 4 book chapters, over 25 refereed international journal papers and 48 peer reviewed international conference papers. He is a reviewer of several international journal of repute.