European Journal of Operational Research 294 (2021) 831–842

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Distributional regression for demand forecasting in e-grocery

Matthias Ulrich[a,*], Hermann Jahnke[a], Roland Langrock[a], Robert Pesch[b], Robin Senge[b]

[a] Department of Business Administration and Economics, Bielefeld University, Universitätsstraße 25, Bielefeld 33615, Germany
[b] inovex GmbH, Ludwig-Erhard-Allee 6, Karlsruhe 76131, Germany

ARTICLE INFO

Article history:
Received 20 December 2018
Accepted 13 November 2019
Available online 20 November 2019

Keywords:
Forecasting
Inventory
E-commerce
Retailing

ABSTRACT

E-grocery offers customers an alternative to traditional store grocery retailing. Customers select e-grocery for convenience, making use of the home delivery at a selected time slot. In contrast to store retailing, in e-grocery in-stock information for stock keeping units (SKUs) becomes transparent to the customer before substantial shopping effort has been invested, thus reducing the personal cost of switching to another supplier. As a consequence, in-stock availability of SKUs has a particularly strong impact on the customer's order decision, resulting in higher strategic service level targets for the e-grocery retailer. To account for these high service level targets, we propose a suitable model for accurately predicting the extreme right tail of the demand distribution, rather than providing point forecasts of its mean. Specifically, we propose the application of distributional regression methods — so-called Generalised Additive Models for Location, Scale and Shape (GAMLSS) — to arrive at the cost-minimising solution according to the newsvendor model. As benchmark models we consider various regression models as well as popular methods from machine learning. The models are evaluated in a case study, where we compare their out-of-sample predictive performance with respect to the service level provided by the e-grocery retailer analysed.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

In retailing, inventory exceeding or falling short of customer demand generally causes monetary consequences for the retailer. Stock-out situations result in short-term shortage costs due to missed sales, whereas excess inventory generates short-term inventory costs due to spoilage and operational inefficiencies. For each demand period, retailers face the decision problem of selecting the inventory level that minimises the expected total costs. For a given stock keeping unit (SKU), the costs of shortage and excess inventory together determine the (α) service level as the cost-optimal in-stock probability of the SKU. Vice versa, any given service level corresponds to a particular ratio of shortage and excess inventory costs. For example, operating with a service level target of say 90% — which would be optimal if shortage costs are nine times higher than excess inventory costs — retailers select inventory levels such as to completely fulfil customer demand in 90% of all demand periods. However, due to high opportunity costs in retail practice, a service level selected exclusively based on short-term costs may not make strategic sense. Instead, long-run strategic objectives such as customer loyalty and market growth based on customer satisfaction usually have the most substantial impact on the service-level decision (Anderson, Fitzsimons, & Simester, 2006). Within the fast-expanding electronic grocery (e-grocery) market — which is an alternative to traditional store grocery retailing — the service level is particularly relevant for inventory management. In e-grocery, customers order groceries online and the retailer delivers the purchase to the household or company. Customers use e-grocery for convenience, making use of home delivery at a selected time so that no visit to a physical store is required. However, if customers are dissatisfied with in-stock availability, then there is a particularly high risk that they cancel the online shopping process, and perhaps even refrain from ordering online in the future. This results from the fact that the convenience of online shopping is reduced once customers need to place a second order or visit a physical store to buy products affected by stock-outs. In-stock availability can hence be expected to be of crucial importance regarding the customers' order decision. Due to the increased risk of e-grocery customers cancelling entire orders as a consequence of stock-outs, shortage costs in this market are substantially higher than inventory costs. For the e-grocery retailer considered as a case study in this work, an average shopping basket is worth $\sim 80-90$ EUR, whereas spoilage generates average costs of about 1.20 EUR per unit, which further underlines the importance of avoiding order

* Corresponding author.
E-mail addresses: matthias.ulrich@uni-bielefeld.de (M. Ulrich), hjahnke@uni-bielefeld.de (H. Jahnke), roland.langrock@uni-bielefeld.de (R. Langrock), robert.pesch@inovex.de (R. Pesch), robin.senge@inovex.de (R. Senge).

https://doi.org/10.1016/j.ejor.2019.11.029
0377-2217/© 2019 Elsevier B.V. All rights reserved.

cancellations. As a consequence, e-grocery retailers operate with very high strategic service level targets, e.g. 97–99% in the case of the retailer analysed in our case study, which is in line with service level targets of other international e-grocery retailers, such as Ocado with 98.8% (Ocado Group, 2015).

An additional difference between e-grocery and traditional retailing lies in the type of information available on customer demand. In traditional store retailing, information on customer demand typically results from point-of-sale data. These data often do not reveal the true demand preferences due to stock-outs affecting the individual purchase, and are therefore effectively censored. This distortion in customer demand for any given SKU also affects demand data for complementary SKUs that are not purchased although available (e.g. mozzarella, if tomatoes are out), and demand data of selected substitutes that generate sales without original demand. Predicted demand of an SKU that was subject to stock-outs in the past and of corresponding complementary SKUs will therefore be potentially too low, whereas predicted future demand of selected substitutes will potentially be too high (Anupindi, Dada, & Gupta, 1998). In contrast to traditional retailing, the customer's online ordering process of the e-grocery retailer considered in this work allows for the monitoring of customer preferences *before* stock-out information becomes available to the buyer and, therefore, yields uncensored demand data. Stock-outs do affect sales data but not the demand data revealed by the customer's online search behaviour, which precedes the final purchase decision. Tracing the customer's search behaviour hence yields uncensored demand data. Moreover, in e-grocery a customer selects a future delivery time slot with up to fourteen days in advance to make sure he or she is able to receive the order at home. Depending on the delivery time slot selected by the customer, the e-grocery retailer can include the customer demand information in the replenishment order. Thus, the e-grocery retailer does not have to hold all inventory requested at the time of the customer order, but may replenish the SKU before the order needs to be dispatched.

The business problem analysed in this paper requires customer demand forecasts for each SKU in each local distribution warehouse, a so-called fulfilment centre (FC), for each demand period, which is the lowest hierarchical level in retail demand forecasting. At this level of detail, many different characteristics may affect the demand, rendering demand forecasting very challenging (Fildes, Ma, & Kolassa, 2018). Fig. 1 illustrates a small subset of the data considered in our case study, namely customer demand for the SKU grapes in the months May to August 2017 for FC 1 (of six FCs analysed in our case study). Here one observation equals one demand period t, i.e. one day of delivery. We find that demand is highly variable, but with recurring peaks on Mondays. These peaks are due to the increased proportion of business-to-business transactions in e-grocery relative to store retailing. Businesses typically make use of the grocery delivery on Mondays to supply their employees or guests with fruits, coffee and complementary SKUs such as milk.

Given the wide range of demand patterns of SKUs on offer (e.g. fast-moving or slow-moving SKUs with regular or irregular demand), it seems unlikely that a simple statistical modelling framework such as linear regression will be suitable across the range of SKUs. For demand modelling, most results reported in the literature do indeed identify potential improvements by using nonlinear models (Steiner, Brezger, & Belitz, 2007, Lang, Steiner, Weber, & Wechselberger, 2015, Weber, Steiner, & Lang, 2017). Furthermore, since we are interested in very high service levels, and hence need to be able to accurately predict the extreme right tail of the forecast distribution, any regression method that focuses on the *mean* would not seem to be an intuitive choice. Distributional forecasting incorporates demand uncertainty by estimating predictive densities. The difference between the estimated quantile and the estimated mean can be interpreted as the safety stock level, resulting from imperfect demand observations (Taylor, 2007, Fildes et al., 2018). Quantile regression, where specific quantiles of the response variable (demand) are linked to covariates, constitutes an obvious approach given the high service-level targets (Kneib, 2013). However, using quantile regression to predict the extreme (right) tail of the response distribution can be problematic, since the corresponding parameter estimators can be highly imprecise due to data scarcity in the extreme tails (Hohberg, Peter, & Kneib, 2018). In recent years, distributional regression methods, which allow flexible modelling of covariate effects on *any* of the distributional parameters (including mean and variance), have rapidly gained popularity as versatile statistical modelling tools that allow consideration of various aspects of the response distribution (cf. Mayr, Fenske, Hofner, Kneib, & Schmid, 2012, Klein, Kneib, Lang, & Sohn, 2015). In contrast to quantile regression methods, which are nonparametric, distributional regression methods are parametric. Sachs and Minner (2014) compared parametric and nonparametric modelling approaches for estimating different censoring levels using data from a large European retail chain. They showed that their nonparametric model copes well with highly censored data, but also pointed out that the parametric alternatives performed best when there was little censoring in the demand.

Considering data from a leading German omni-channel retailer, here we aim to exploit the new types of data in e-grocery that are not available in traditional store retailing, i.e. uncensored customer demand and partly known future demand data, to model the entire demand distribution. We propose the application of Generalized Additive Models for Location, Scale and Shape (GAMLSS) for building flexible distributional regression models to forecast demand in e-grocery retailing. In particular, these models consider not only the mean of future demand, but also its variance and potentially even the shape as functions of covariates. In addition, the GAMLSS class allows a flexible choice of the distributional family assumed for the response variable demand. Thus, GAMLSS allows us to tailor the regression model to whatever complex pattern we find, while likely being more robust than quantile regression methods. We compare the performance of GAMLSS with basic regression as well as quantile regression models, evaluating the models by comparing their costs realised out-of-sample, for a given service level with the corresponding costs for shortage and excess inventory, respectively. Given the increasing popularity of machine learning methods (see, e.g., Carbonneau, Laframboise, & Vahidov, 2008, Ferreira, Lee, & Simchi-Levi, 2016), we also include random forests (Breiman, 2001) and quantile regression forests (Meinshausen, 2006) as additional benchmarks. For the SKUs considered in our case study, we find that models from the GAMLSS class tend to outperform the benchmarks, with the (cost-)optimal distributional assumption to be made for the response variable varying across SKUs.

2. Problem statement and motivation

2.1. Demand forecasting and the newsvendor problem

The e-grocery retailer considered in our case study offers a large number of perishable SKUs in the product category fruits, vegetables, and meat. Internal quality requirements of the retailer restrict the shelf life of these SKUs to one demand period. In addition, SKUs with stock-outs in the demand period are not delivered to the customer at a later demand period. We hence assume that the customer demand and the sales period are identical for the SKUs analysed in the case study. As a result, excess inventory cannot be sold in the following demand period and thus generates spoilage.

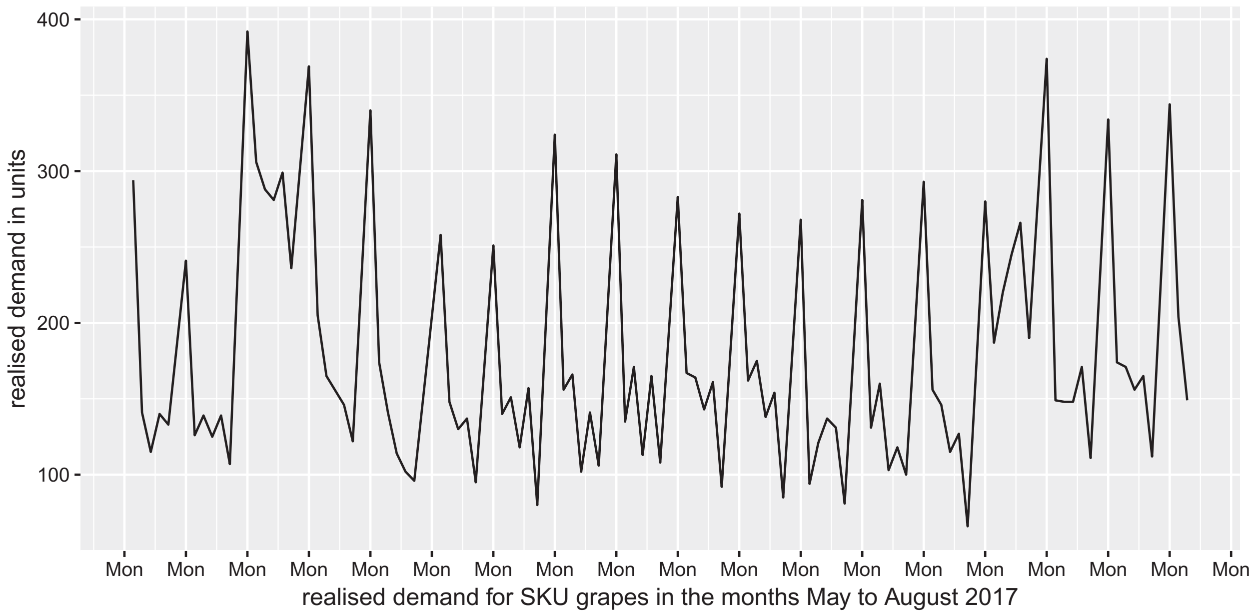

Fig. 1. Realised demand for the SKU grapes in units (package of 500 gr), with highly variable demand but recurring peaks on Monday for FC 1 (of six FCs analysed in our case study).

To capture the asymmetric economic impact of absolute forecasting errors for each demand period t, we introduce the total cost C_t resulting from any potential mismatch between inventory level and realised demand. In demand period t, each excess unit of inventory generates a cost of h, while each unit we fall short of customer demand generates a cost of b. Furthermore, we use D_t to denote the stochastic customer demand. We then aim at minimising the expected total cost,

$$E[C_t(q_t)] = hE(q_t - D_t)^+ + bE(D_t - q_t)^+,$$

with respect to the inventory level at the beginning of the demand period, q_t. The optimal q_t defines the corresponding replenishment order quantity of the retailer.

For single and independent demand periods with stochastic customer demand, the newsvendor problem provides the solution to the optimisation problem above. The newsvendor problem is one of the classical applications in the literature on inventory management for problems with characteristics as given here (Zipkin, 2000). The optimal inventory level is obtained as

$$q_t^* = \underset{q_t}{\operatorname{argmin}}\, E[C_t(q_t)] = F_t^{-1}\big(b/(b+h)\big), \quad (1)$$

where F_t is the (true) cumulative distribution function of the demand distribution in period t, and $b/(b+h)$ is the optimal demand quantile given b and h. The ratio $b/(b+h)$ can also be interpreted as the inventory service level selected by the retailer. In practice, the optimal solution to the newsvendor problem given in (1) is not available since the cumulative distribution function F_t describing the stochastic demand is unknown. However, we can use data collected before time t to statistically model realised demand as a function of features (e.g. known demand at the time of the replenishment order), and subsequently derive q_t using the estimated distribution function $\hat{F}_t$ obtained under the model. Thus, we are looking for a suitable distributional regression model, rather than point forecasts only, to minimise the costs of the e-grocery retailer.

2.2. Demand forecasting in the existing literature

In the retailing context, the newsvendor problem is one of the most intensively studied inventory management problems in literature; see Khouja (1999) and Qin, Wang, Vakharia, Chen, and Seref (2011) for extensive literature reviews. Early papers from this research area assume that the demand distribution is completely known (Arrow, Harris, & Marschak, 1951). Scarf, Arrow, and Karlin (1959) relaxed this assumption by assuming that only the mean and the standard deviation are known. Since those early contributions, various parametric (e.g. Nahmias, 1994, Agrawal & Smith, 1996) and nonparametric (e.g. Lau & Lau, 1996, Godfrey & Powell, 2001) approaches have been proposed to identify the cost-optimal inventory level q for the case where the demand distribution is unknown and hence needs to be estimated or otherwise specified. Parametric approaches to determine the optimal order quantity focus on the estimation of the mean demand by a point prediction together with an estimate of the standard deviation. Based on the expected demand pattern, different distributional assumptions have been proposed, e.g. normal (Nahmias, 1994), gamma (Burgin, 1975), Poisson (Conrad, 1976), and negative binomial (Agrawal & Smith, 1996). In retail practice with $b > h$, the inventory level q minus the estimated mean is often interpreted as the safety stock to cope with forecasting errors (e.g. Baker, Magazine, & Nuttle, 1986).

Linear regression is the most established approach for modelling directed relationships and as such still is a popular benchmark model also in demand forecasting. However, linear regression involves several restrictive assumptions, namely linearity of the predictor, homoscedasticity and, when forecast distributions are of interest, also normally distributed errors. For some of the SKUs on offer, these will be violated, for example if the distribution of slow-moving SKUs is neither normal nor symmetric (Ramaekers & Janssens, 2008). In those instances, demand forecasts obtained through linear regression may be imprecise. Indeed, in recent years, several contributions have reported improved forecasts when using nonlinear regression models (see, e.g., Steiner et al., 2007, Lang et al., 2015, and Weber et al., 2017).

Quantile regression is a natural nonparametric alternative which focuses on forecasting selected quantiles rather than the mean (Koenker & Hallock, 2001, Maciejowska, Nowotarski, & Weron, 2016). In particular, within quantile regression we avoid making any distributional assumption for the response. Haupt, Kagerer, and Steiner (2014) found that quantile regression, with P-spline smoothers for estimating potentially nonlinear covariate

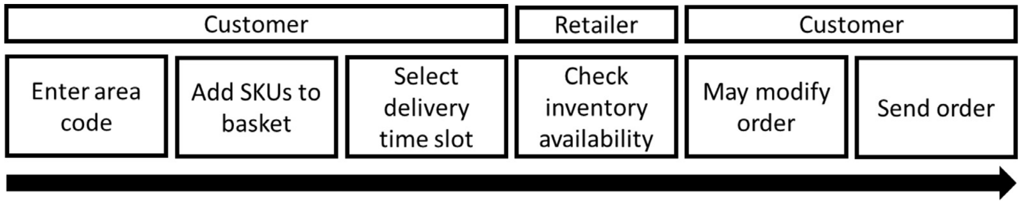

Fig. 2. Customer order process in e-grocery.

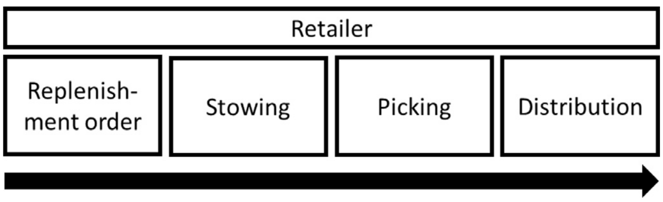

Fig. 3. Replenishment and fulfilment process in e-grocery.

effects, outperformed parametric alternatives for sales predictions. However, a potential problem with using quantile regression methods for e-grocery demand forecasting is that we often see notable peaks in the realised demand data, to which quantile methods may react more strongly than distributional regression models (cf. Hohberg et al., 2018). These peaks are due, *inter alia*, to the high proportion of business-to-business transactions. Meinshausen (2006) proposed quantile regression forests as a potentially more robust alternative to quantile regression.

Finally, for point forecasts of the mean, various machine learning techniques have rapidly increased in popularity as an alternative to linear regression (see, e.g., Carbonneau et al., 2008, Ferreira et al., 2016). For example, random forests have been successfully applied in the context of load forecasting on the electricity market (Lahouar & Ben Hadj Slama, 2015).

2.3. E-grocery business processes

Irrespective of the particular approach taken, we aim to make use of the new types of data in e-grocery that are not available in traditional retailing. To better understand the corresponding data, we first introduce the business processes that generate these new types of data. The complete process of a customer's order is displayed in Fig. 2. Customers first enter the area code, which determines the associated FC with a given assortment. During the shopping process, no inventory information is provided to the customer, such that any desired SKU can simply be added to the basket, without restrictions resulting from its availability status. This allows for the observation of customer preferences *before* stock-out information is made available. At the customer checkout, an algorithm checks the inventory availability of the requested SKUs for the selected delivery time slot. In case of missing inventory, a suitable substitute is offered to the customer, who can then modify the order.

The replenishment and fulfilment process of the e-grocery retailer is shown in Fig. 3. The national and regional distribution centres supply the FC based on replenishment orders. Operational fulfilment processes include stowing, picking, and the distribution to the customer. It typically takes three days from placing the replenishment order to receiving the goods at the FC. Thus, if the customer orders more than three days in advance to the delivery slot, then known demand can be considered for the replenishment order.

2.4. Exploratory analysis of the e-grocery data

In the following, we explore the e-grocery data available in our case study in order to motivate the choice of the class of GAMLSS. To illustrate some of the key patterns, Fig. 4 displays the relationship between selected explanatory variables (features) and the response variable, realised demand, for the SKU grapes from September 2015 to August 2017, for FC 1. We find, *inter alia*, the following patterns:

1. *Nonlinearity* — Fig. 4(a) shows the relationship between known customer demand at the replenishment decision time and realised demand.[1] Realised demand equals or exceeds known demand for each observation. For relatively low known demand, i.e. below 100 units, the size of additional demand occurring during the replenishment period is relatively high compared to situations where known demand is already high. This indicates that the functional relationship between realised demand and known demand is nonlinear.
2. *Heteroscedasticity* — Fig. 4(b) relates the realised demand to the feature 'median demand of the same weekday in the previous month', showing that the variance in demand increases with increasing values of this feature.
3. *Skewness* — Fig. 4(c) shows positive skewness in the distribution of the realised demand. The upper whisker and the 0.75 quantile are farther from the median than the 0.25 quantile and the lower whisker.

3. Distributional regression and benchmarks

3.1. GAMLSS

Given the complex patterns found in the data, we propose to use Generalised Additive Models for Location, Scale and Shape (GAMLSS), as they allow a flexible selection of distributions for the demand, and also a flexible modelling of covariate effects on any of the distributional parameters (Rigby & Stasinopoulos, 2005). The GAMLSS class is an extension of Generalised Linear Models (Nelder & Wedderburn, 1972) and Generalised Additive Models (Hastie & Tibshirani, 1986).

In GAMLSS, a parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)$ — rather than the mean only — of the response variable's probability (density) function $f(y|\boldsymbol{\theta})$ is modelled as a function of covariates. The p parameters being modelled determine the location, scale and shape of the distribution, with the value of p varying across the different types of distributions that can be assumed for the response. More specifically, it is assumed that the observations y_i, $i = 1, 2, \ldots, n$, are independent, each with an associated parameter vector $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, ..., \theta_{pi})$ and probability (density) function $f(y_i|\boldsymbol{\theta}^i)$. In our

[1] Here the feature 'known demand' is the customer demand information for the corresponding demand period that is already available at the replenishment decision time due the customer's option to order with up to fourteen days in advance. This demand information can be included as a feature to estimate realised demand. Realised demand equals the monitored customer preferences before stock-out information becomes known to the buyer.

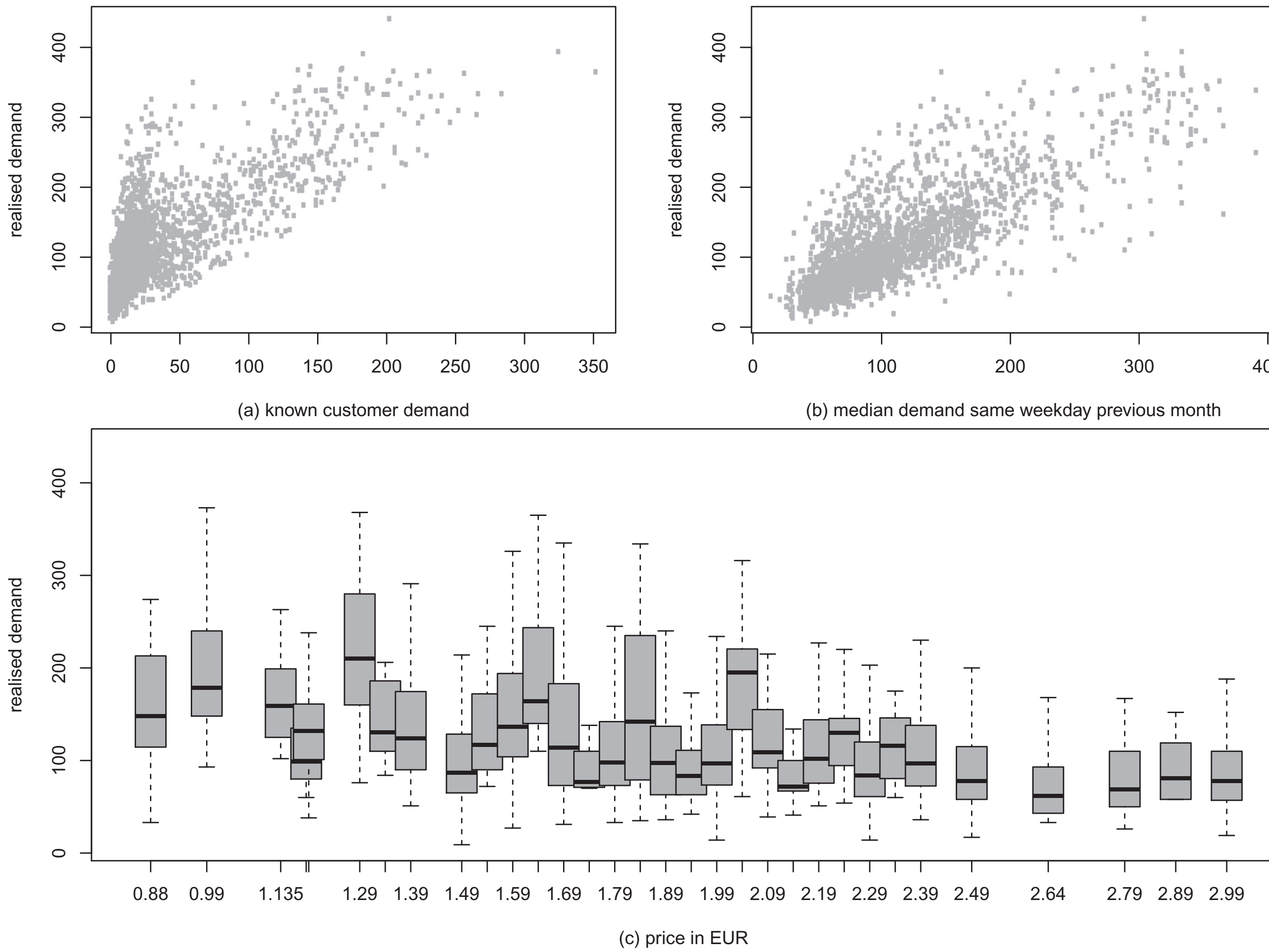

Fig. 4. The relationship between selected explanatory variables (features) and the response variable, realised demand, for the SKU grapes within the months September 2015 to August 2017 for FC 1. In (c), boxplots of realised demand are shown for each price level in the data available.

case, each observation y_i corresponds to the realised demand for some SKU on a given day, and the covariates used to explain y_i will be features built from data on demand prior to the day considered (as detailed below). We adopt the notation from Rigby and Stasinopoulos (2005), using $\boldsymbol{\theta}_k = (\theta_{k1}, \theta_{k2}, \dots, \theta_{kn})$ to denote the vector of the k-th distributional parameter being modelled (one for each daily observation of demand). For $k = 1, \dots, p$, a known monotonic link function $g_k(\cdot)$ then relates $\boldsymbol{\theta}_k$ to features and random effects through an additive model,

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \boldsymbol{X}_k\beta_k + \sum_{j=1}^{J_k} \boldsymbol{Z}_{jk}\gamma_{jk}, \tag{2}$$

where g_k is applied componentwise, $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are vectors of length n, $\boldsymbol{X}_k$ is a known design matrix of dimension $n \times J'_k$, $\beta_k = (\beta_{1k}, \beta_{2k}, ..., \beta_{J'_k})^T$ is a parameter vector of length J'_k, $\boldsymbol{Z}_{jk}$ is a fixed known $n \times \upsilon_{jk}$ design matrix and γ_{jk} is a υ_{jk}–dimensional random variable. Thus, for $k = 1, \dots, p$, in the most general case the linear predictor η_k includes the parametric component $\boldsymbol{X}_k\beta_k$ and in addition additive components $\boldsymbol{Z}_{jk}\gamma_{jk}$. Rigby and Stasinopoulos (2005) call model (2) the GAMLSS. The first two distributional parameters θ_1 and θ_2 are typically characterised as location and scale parameter, denoted by μ and σ, respectively, whereas the remaining parameters are generally characterised as shape parameters.

For $\boldsymbol{Z}_{jk} = \boldsymbol{I}_n$, where $\boldsymbol{I}_n$ is an $n \times n$ identity matrix, and $\gamma_{jk} = \boldsymbol{h}_{jk} = h_{jk}(\boldsymbol{x}_{jk})$ for all combinations of j and k, the GAMLSS model formulation (2) is semi-parametric given by

$$g_k(\boldsymbol{\theta}_k) = \eta_k = \boldsymbol{X}_k\beta_k + \sum_{j=1}^{J_k} h_{jk}(\boldsymbol{x}_{jk}), \tag{3}$$

where the $\boldsymbol{x}_{jk}$ are vectors of length n, for $j = 1, 2, \dots, J_k$ and $k = 1, 2, \dots, p$. The function h_{jk} is an unknown function that is componentwise evaluated for the feature vector $\boldsymbol{x}_{jk}$. The explanatory vectors $\boldsymbol{x}_{jk}$ are assumed to be known (Rigby & Stasinopoulos, 2005). In our case study, we implement a P-spline smoother to account for potential nonlinear relationships between features and realised demand.

The GAMLSS allows to make various distributional assumptions for the response; see Rigby and Stasinopoulos (2005). For demand forecasting, the normal (Nahmias, 1994), gamma (Burgin, 1975), Poisson (Conrad, 1976), and negative binomial distribution (Agrawal & Smith, 1996) are the most established distributions in the literature. As a consequence, these are the distributions that we consider in our case study (with the corresponding models labeled as GAMLSS_normal, GAMLSS_gamma, GAMLSS_Poisson and GAMLSS_negBin, respectively). GAMLSS offers two negative binomial distributions, which differ in the definition of the variance. Type I defines the variance by $(\mu + \sigma\mu^2)$ and type II by $(\mu + \sigma\mu)$,

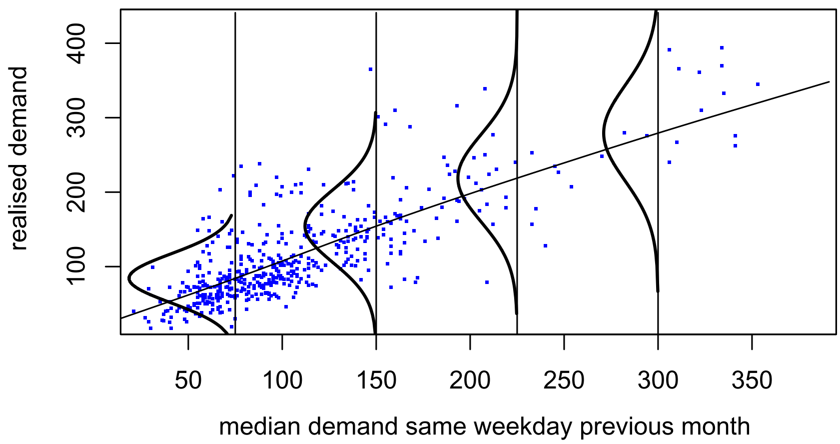

Fig. 5. Realised demand in FC 1 for grapes as a function of median demand on the same weekday over the previous month, together with the fitted GAMLSS, including the error distribution at selected values of the covariate.

given a mean μ and a standard deviation σ. For any given distributional assumption, and assuming independence of the individual realised demand values, the likelihood function of a GAMLSS is easily obtained. To fit a given GAMLSS to demand data using maximum likelihood, we use the R package `gamlss`.

To fix these ideas, in Figure 5 we display a very simple GAMLSS, with normally distributed response, fitted to the demand data collected on the SKU grapes from September 2015 to August 2017 for FC 1, as shown already in Figure 4. More specifically, the model that was fitted is specified as follows:

$$y_i \sim N(\mu_i, \sigma_i^2), \quad \mu_i = h_1(x_i), \quad \log(\sigma_i) = h_2(x_i),$$

where y_i is the i–th realised demand and x_i is the median demand on the same weekday, calculated for grapes in FC 1. As an example, we use median customer demand of the previous five Mondays as a feature when estimating customer demand for the following Monday. The functions h_1 and h_2 were estimated nonparametrically using P-splines. For notational simplicity, some indices were dropped here compared to the general model formulation in Eq. (2).

The model displayed in Fig. 5, while simplistic, adequately addresses the heteroscedasticity shown in Fig. 4 by estimating the variance σ_i^2 to increase for increasing covariate values. As a consequence, under such a model formulation, the cost-optimal inventory level at high covariate values would be farther away from the regression line than for small covariate values, thereby accounting for the increased uncertainty regarding customer demand. This pattern in the choice of the inventory level is in line with the theoretical results by Song (1994), who by means of stochastic comparison methods showed the impact of growing demand variability on optimal inventory levels and the corresponding economic costs.

3.2. Benchmark models

Based on our literature review in Section 2.2, we consider linear regression (Reg), log-linear regression (Log-linear), log-log regression (Log-log), quantile regression (QuantReg), random forests (RanForest), and quantile regression forests (QuantRegForest), as performance benchmarks for the GAMLSS approach proposed here.

When using a linear regression model, we predict the *conditional mean* of the response distribution and, assuming normality of the residuals, derive the target quantile based on the estimated constant error variance. The log-linear model expresses the logarithm of the demand as a function of a linear predictor comprising the covariates as they stand. Within the log-log model, the covariate values are also log-transformed before modelling, corresponding to a multiplicative model. Quantile regression methods directly predict any *conditional quantile* of the response distribution (Koenker & Hallock, 2001). For fairness of comparison with the GAMLSS approach, we follow Haupt et al. (2014) and estimate covariate effects within quantile regression nonparametrically using P-splines, as we do within GAMLSS. Random forests represent an ensemble learning technique that combines predictions from a specified number of tree predictors. Each tree is constructed independently by using a sub-sample of all observations available (Breiman, 2001). The output is again a prediction of the *conditional mean* of the response variable. As a point forecast only, i.e. without accompanying distributional assumption for the residuals (such as the normal in case of basic linear regression), this prediction does not readily allow us to arrive at say the 0.97 quantile. In order to be able to compare the random forest-based prediction to the approaches that do provide us with quantiles, we thus complement the point prediction with an additional distributional assumption. Specifically, we assume normality of the residuals, with a constant plug-in standard deviation as estimated from one year of training data. We tested alternative distributional assumptions for the residuals — gamma, Poisson, and negative binomial — but none of these led to notably better results, such that we do not report them here for conciseness. Finally, quantile regression forests constitute an alternative to random forests that again allow the direct estimation of any *conditional quantile* (Meinshausen, 2006).

The non-standard models considered are implemented using the following R packages: *qgam* (Matteo & Wood, 2019) for smooth additive quantile regressions, *quantregForest* (Meinshausen, 2006) for the quantile regression forest, *randomForest* (Breiman, Cutler, Liaw, & Wiener, 2018) for random forests, and *gamlss* (Rigby & Stasinopoulos, 2005) for GAMLSS.

4. Model training, forecasting and performance evaluation

4.1. Feature engineering and selection

The demand distribution F_t is estimated using features (i.e. explanatory variables). Feature engineering describes the process of generating suitable features from the data available. Both the general pattern of the demand distribution as well as time series effects are taken into account by considering historic demand quantiles (5%, 50% and 95%), building corresponding features using data from a) the previous quarter, b) the previous month, and c) the previous two weeks (Lu, 2014, Kawamura, Nomoto, & Kuo, 2015). We consider 12 features in total, also including price and known demand extracted directly from the raw data:

(1) demand on the same weekday of the previous week in units;
(2) demand on the same weekday of the second to last week in units;
(3) median demand on the same weekday over the previous month in units;
(4) median demand on the same weekday over the previous quarter in units;
(5) median demand over the previous month in units;
(6) median demand over the previous quarter in units;
(7) 95% demand quantile over the previous month in units;
(8) 95% demand quantile over the previous quarter in units;
(9) 5% demand quantile over the previous month in units;
(10) 5% demand quantile over the previous quarter in units;
(11) price;
(12) known customer demand in units.

In principle, additional features such as cross prices and marketing events might also improve the models' predictive accuracy. However, for this case study, within which we choose to focus on comparing competing modelling frameworks (as opposed to optimising the performance of any individual model specification), we restricted our case study to price and demand data of the five SKUs presented.

The random forest-based approach, in contrast to all other methods implemented, implicitly conducts feature selection within the training stage (see, e.g., Ludwig, Feuerriegel, & Neumann, 2015). The other methods may benefit from reducing the set of features considered, in particular to avoid overfitting, but also to prevent problems arising from multicollinearity. We tested three different approaches to feature selection, in each case applying the given procedure within each of the methods considered:

(i) no feature selection, i.e. all features are included (for each SKU);
(ii) for each individual SKU, features are selected by a stepwise forward selection based on (a) the Akaike Information Criterion (AIC) or (b) the Bayesian Information Criterion (BIC);
(iii) first, features are pre-selected based on multicollinearity statistics and taking into account business experience, then for each SKU features from the pre-selected set are chosen on (a) an AIC-based or (b) a BIC-based stepwise forward selection.

Within (iii), we pre-selected the following features: median demand on the same weekday over the previous month (3), median demand over the previous quarter (6), price (11), and known customer demand (12). This combination of features provides a good balance between capturing relevant economic information on the one hand — (3) is a proxy variable for capturing day-of-week effects, (6) captures trends in the fast-growing e-grocery business, and the role of (11) and (12) is obvious — yet limiting multicollinearity on the other hand. In addition, all of these four variables are easy to develop and interpret in retail practice. All pre-selected features were retained by the AIC for all SKUs except for meat, where price was excluded by the AIC. When using the BIC, price was excluded also for the SKU tomatoes. Approach (iii) performed best in terms of realised costs over the validation period. Overall, the much more parsimonious models obtained under (iii) yielded total costs 8% lower on average (both when using AIC and also when using BIC) than the costs obtained under the full models, with approach (i). Feature selection based on the AIC only, using approach (ii) above, led to roughly the same total costs as when using the full model, (i), which is not too surprising in view of the generosity of the AIC with regard to model complexity. Method (ii) using the BIC provides more parsimonious models and lower cost as compared to when using the AIC, but still higher costs as compared to when method (iii) is used. We also tested an alternative, more formal pre-selection using variance inflation factors, which however led to an implausible selection of features and overall higher costs even than (i). Our findings demonstrate the trade-off between purely relying on formal statistical criteria like VIF on the one hand and bringing in economic rationale for covariate selection on the other hand. In the following, we show the results obtained with (iii) using the AIC, giving the results obtained under (i), (ii), and (iii) using the BIC in the online supplementary material. We decided to focus primarily on the AIC, as this criterion aims at selecting the model with the highest predictive accuracy — which is in line with our overall objective — whereas the latter aims at selecting the model most likely to be true (see, e.g., Shmueli, 2010). In retail practice, it may however also be of interest to identify the actual drivers of demand, in which case the BIC can be preferable.

4.2. Training data

Our data set covers demand periods from September 2015 to August 2017 and six different e-grocery FCs. We consider daily data, i.e. each demand period t refers to one day of delivery. For demand forecasting and hence validation of the different approaches considered, we use all demand observations from September 2016 to August 2017. We test five different SKUs within the SKU-category fruits, vegetables, and meat, namely tomatoes, carrots, grapes, mushrooms, and minced meat. All of these SKUs were listed for the entire period, exhibiting price changes throughout (though with only little variability in case of the SKU minced meat).

We train each model based on twelve months of data to subsequently forecast demand in the following month. For example, we train each model based on data from September 2015 to August 2016 to forecast demand in September 2016. With this sliding-window approach, we always train on the most recent twelve months of data. In our models, we do not include variables explicitly addressing seasonality. We do however accommodate, at least to some extent, seasonal effects by means of features that indicate the current (seasonal) level of demand, including for example the median demand in the previous month, but also the price, which in retail practice reflects the supply-side seasonality over the year. For each demand period t and each of the parametric models considered, we obtain an estimate $\hat{F}_t$ for the demand distribution F_t, which we apply to derive q_t according to the newsvendor problem, for any given demand quantile. For the nonparametric models that we consider — quantile regression and quantile regression forests — we derive q_t for selected demand quantiles directly from the data, i.e. without previously having to build an estimate $\hat{F}_t$. Thus, for each SKU we derive demand distributions for about 300 demand periods within the validation period, which results in >350 K predictions of q_t when considering 14 models and 99 possible service levels between 0 and 100.

4.3. Out-of-sample prediction of demand

Figure 6 illustrates the prediction step for the SKU grapes, showing the optimal q_t given a target service level of 97%, as obtained under the GAMLSS models, and additionally the realised demand, for the 23 demand periods in April 2017 observed in FC 1. Since we derived the optimal inventory level q_t at the 97% service level target, q_t should equal or exceed realised demand in 97% of all observations. But for our small subset of example prediction data observed for FC 1, modelling via the GAMLSS_Poisson model yielded an inventory level q_t that is below realised demand for three demand periods (01–04, 03–04, and 20–04), corresponding to a service level of only 87% (for this month). All

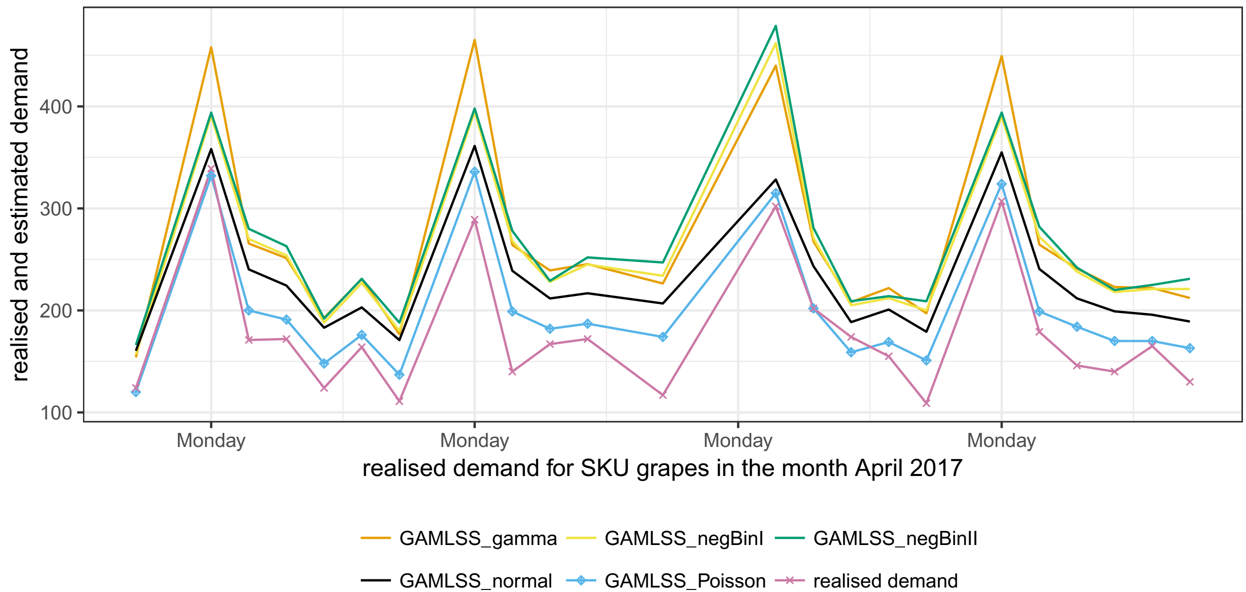

Fig. 6. Predicted 0.97 demand quantiles as obtained with the different GAMLSS models, and realised demand for the SKU grapes, in April 2017 for FC 1.

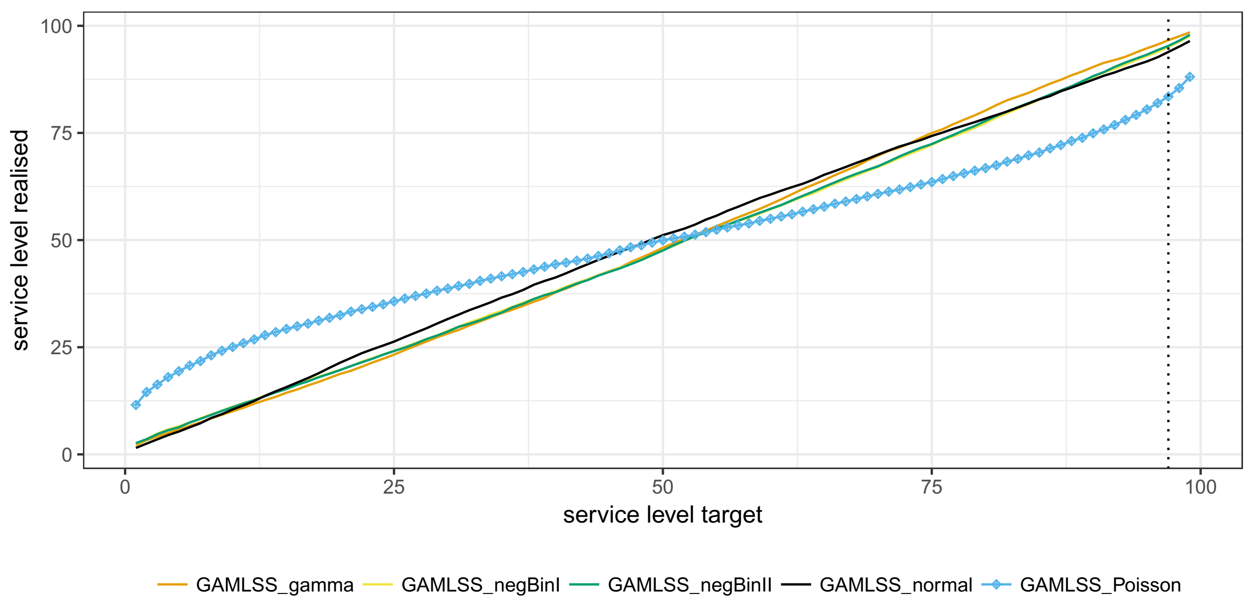

Fig. 7. Service level target vs. realised service level of the GAMLSS models for all SKUs and for all FCs in the validation period September 2016 to August 2017.

other models in this illustrative example yielded inventory levels which in April 2017 never fell short of realised demand. Among these, considering inventory costs for each excess unit of inventory, we prefer inventory levels that are equal to or only slightly above realised demand. In this regard, the GAMLSS_normal model here — i.e. for this subset of the validation data for the SKU grapes — gave the best results, minimising the total costs obtained.

To obtain a more comprehensive picture of the performance of the various GAMLSS formulations, Fig. 7 illustrates the overall deviations obtained under the models between each (integer-valued) service level target and the realised service level, across all SKUs and all six FCs. If a model would meet each service level target exactly, then the corresponding line would follow the 45° line. The plot shows that the GAMLSS_Poisson model leads to substantial deviations for any service level target distinct from 50%, i.e. whenever costs are asymmetric: when the service level target is below 50% (above 50%), then the GAMLSS_Poisson model produces inventory levels that are too high (too low).

In practice, only the service level target of the e-grocery retailer is pertinent for model evaluation. Fig. 8 shows the percentage deviation between the realised service level and the 97% service level target, for each model and across all SKUs and FCs analysed. The models GAMLSS_Poisson and RanForest show deviations above 9%. All other models show promising results, achieving the service level target with deviations below 3%.

4.4. Performance evaluation

Typical point forecast measures as applied in the existing literature, such as the mean average error (MAE) and the mean average percentage error (MAPE), assume a symmetric evaluation of the forecasting error (see, e.g., Makridakis & Winkler, 1983, Carbonneau et al., 2008). They are therefore inadequate to evaluate the

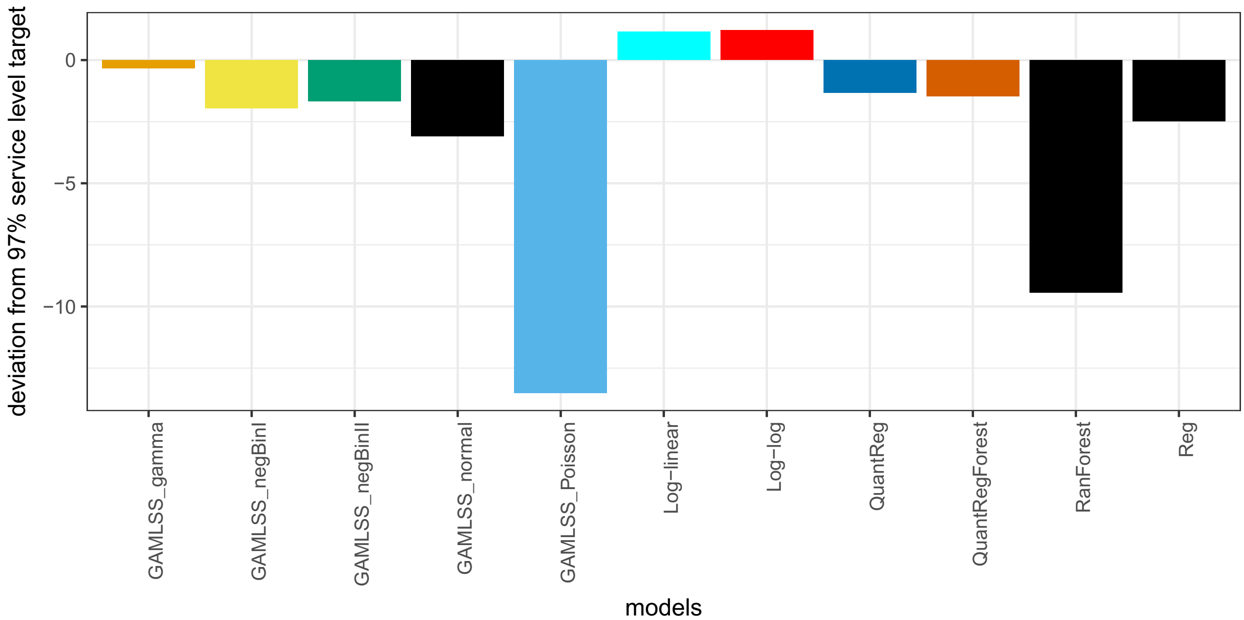

Fig. 8. Model deviation from 97% service level target for all SKUs and for all FCs in the validation period September 2016 to August 2017.

Table 1
Relative overall total costs as obtained under the different approaches, compared to the benchmark linear regression, for five different SKUs across all FCs.

Method	Tomatoes (%)	Carrots (%)	Grapes (%)	Mushrooms (%)	Meat (%)
GAMLSS_gamma	4	−18	−17	−23	2
GAMLSS_negBinI	−4	−13	−15	−21	−6
GAMLSS_negBinII	−6	−14	−16	−24	−6
GAMLSS_normal	−6	−14	−3	−10	−6
GAMLSS_Poisson	35	44	59	12	55
Log-linear	26	−5	9	−22	26
Log-log	10	−15	−7	−29	14
QuantRegForest	−2	−4	−14	17	7
QuantReg	−7	−13	−7	−4	−6
RanForest	39	49	27	49	53

model performance for very high target service levels and hence asymmetric costs. Moreover, the MAPE is undefined for demand realisations of zero, which may occur for SKUs with intermittent demand in grocery retailing; see Kolassa (2016).

Since shortage and inventory costs are asymmetric in our business problem — and we thus have to evaluate the two directions of error asymmetrically — we aim to compare the out-of-sample forecasting error at the empirical service level target elicited from the retailer in our application. While specifying the numerical value of the inventory cost parameter h is relatively straightforward given corresponding information from the retailer's accounting department, this is not the case for the shortage cost parameter b. Especially for relatively novel business models such as the one under consideration (e-grocery), the cost parameter b must reflect the strategic objectives of the retailer and the financial consequences of medium and long term reactions of customers to stock-outs. The precise value of the parameter b hence cannot be stated. Instead, we apply an indirect specification of the shortage cost parameter b via the relation $\alpha = b/(b+h)$ from the newsvendor problem, where α is the service level. Given the empirical value of the service level and an estimation of the inventory cost parameter h based on the retailer's margin and operational costs, we derive the associated shortage cost as $b = h\alpha/(1-\alpha)$. Given b and h, we then calculate the total costs that occur under the q_t obtained from $\hat{F}_t$, for each demand period t in the validation period September 2016 to August 2017:

$$C_t(q_t) = h(q_t - d_t)^+ + b(d_t - q_t)^+.$$

For each model, we sum up the costs for all demand periods t.

The cost values of b and h have a strong impact on the evaluation of the relative model performance. Assuming, as an example, inventory costs of 1 EUR and a service level target of 97%, we obtain shortage costs of 32.3 EUR for each unit of shortage. Thus, a method that overestimates demand by 32 units generates (slightly) lower costs than an alternative method that underestimates demand by one unit.

5. Results

Table 1 shows the percentage differences in total costs (SKU-specific, for each SKU across all six FCs considered), as obtained under the different approaches under the 97% service level target, in each case compared to the benchmark linear regression:

$$\text{perc. diff. in total costs} = 100\left(\frac{C_t(q_t^*)}{C_t(q_t^{\text{Reg}})} - 1\right),$$

where q_t^{Reg} and q_t^* are the optimal inventory levels according to linear regression modelling and the alternative method under consideration, respectively.

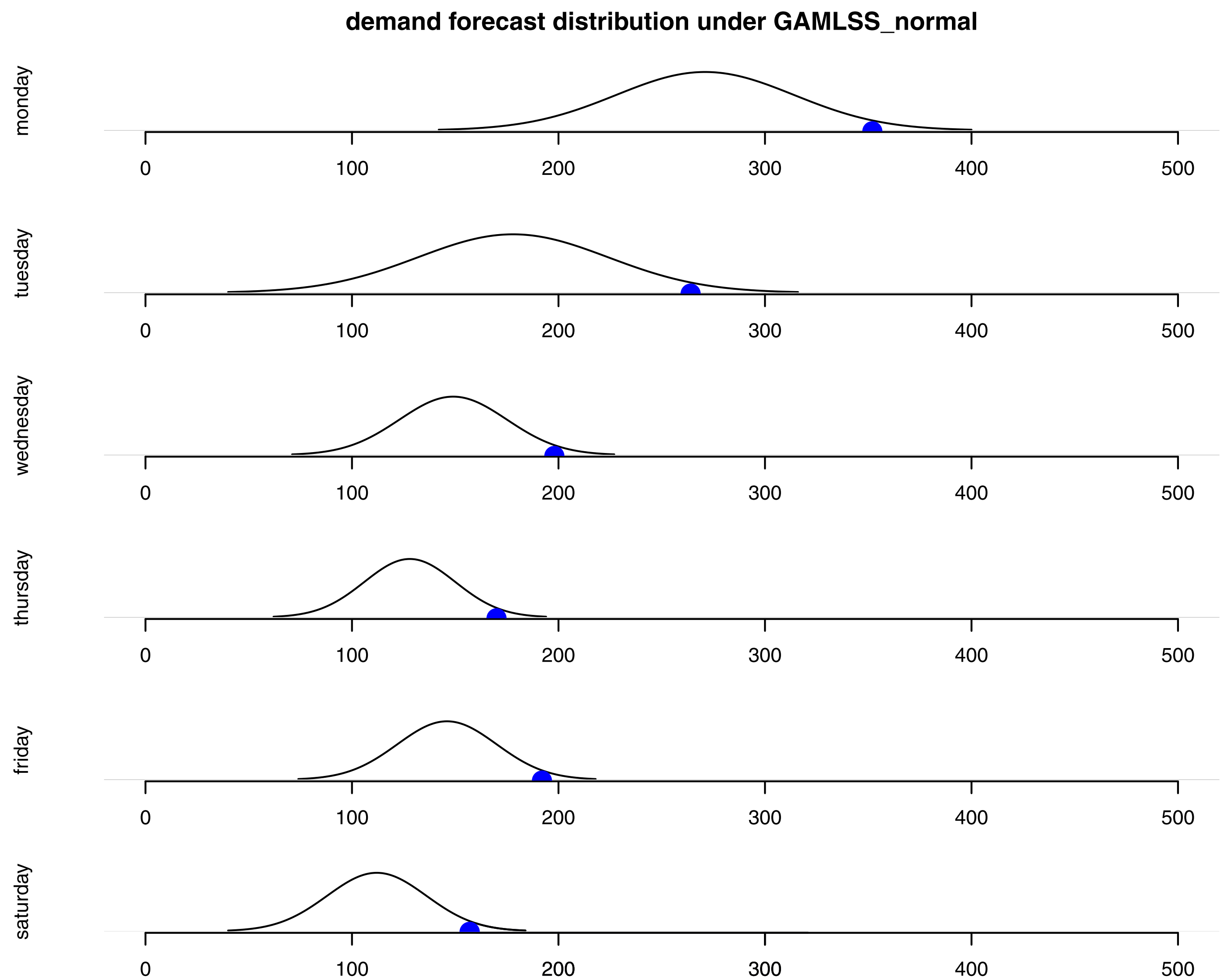

Fig. 9. Demand forecast distribution for the SKU grapes under the model GAMLSS_normal in the week 24-07-2017 to 29-07-2017; the 97% quantiles under each daily forecast distribution are highlighted in blue for FC 1.

While no single model dominates its competitors across all SKUs considered, it is clearly evident that the models from the GAMLSS class — with the exception of GAMLSS_Poisson — consistently performed very well, in many cases yielding either the best (i.e. cost-optimal) or second best performance, and only very rarely performing worse than linear regression (namely in 2 out of 20 cases, and only when using the GAMLSS_gamma formulation). Overall, the GAMLSS_negBinII formulation led to the best results in the given case study in terms of reducing the average percentage costs across all SKUs. Specifically compared to linear regression, we find potential cost reductions of up to 24%. In the online supplementary material to this manuscript, we provide the results of the performance comparison when considering different service level targets, specifically 95% and 90%. As was to be expected, the cost reductions of the distributional regression models compared to linear regression reduce to 17% and 10%, respectively. In other words, the higher the service level target, the more potential there is to reduce overall costs by using distributional regression methods, thereby explicitly addressing heteroscedasticity. However, the relative performance of the different model classes considered does not change across the three service levels considered.

The results indicate that at least one of the assumptions made within basic linear regression, such as homoscedasticity or normality, will typically be violated in this kind of data. The GAMLSS class allows us to tailor the regression model to any given pattern, such that, as expected, in most cases it leads to improved demand forecasts. An exception is the model GAMLSS_Poisson, which with its single-parameter response distribution is actually less flexible than a linear regression model — the implied mean-variance relation of the Poisson here drastically limits the model's flexibility. Overall, the results demonstrate the strengths of GAMLSS by allowing a flexible selection of distributions for the demand, and also a flexible modelling of covariate effects on any of the distributional parameters. However, it is worth noting that there is no distributional assumption within the class of GAMLSS that performs best across all SKUs in our case study. This is perhaps not surprising given the wide range of demand patterns of SKUs on offer.

The multiplicative log-log regression model gave good results for two SKUs considered, but seems to lack stability. The same holds for the log-linear model, which performed poorly for the SKUs tomatoes and meat.

Quantile regression performed only slightly worse than the GAMLSS approaches, and in fact (only just) yielded the lowest out-of-sample costs for two SKUs. The results from Haupt et al. (2014) suggest that the performance of smooth quantile regression could potentially be further improved by imposing monotonicity constraints on the nonparametric estimator of the price effect. However, the same holds for the GAMLSS approach, and in our case study the expected demand was in fact estimated to be monotononically decreasing with increasing price for all but one SKU (carrots). Due to their sensitivity especially for extreme quantiles (e.g. 0.99), quantile regression methods bear the risk of occasionally producing demand forecasts that are way off. In addition, in practice it may not be ideal that these models need to be fitted separately for any target quantile considered. In other words, as distribution-free methods, they do not allow practitioners to select the inventory level based on eyeballing a demand forecast distribution.

For random forests, with our rather ad hoc approach to complement the point forecast by a distributional assumption, the estimated standard deviation from historic data is constant over time. As a consequence, this type of approach cannot accommodate heteroscedasticity, which however is crucially important when trying to capture the extreme (right) tail of the demand distribution. In our example, the performance of the random forest-based approach is indeed very poor, as was to be expected given the systematic deviation from the service level target (see Fig. 8). The performance could potentially be much improved by using alternative, more sophisticated strategies to complement the point forecast in order to obtain the quantile of interest. However, it seems clear that the focus on producing point forecasts will always be a disadvantage in this context.

The retail partner of our case study emphasised the importance of interpretability. In retail practice, the interpretation of demand forecasts, and how they were obtained, is a relevant requirement of the operational management since it allows the inclusion of expert knowledge for forecast modifications in case of exceptional circumstances (cf. Trusov, Bodapati, & Cooper, 2006 Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009, Davydenko & Fildes, 2013). In particular, machine learning techniques are often difficult to interpret. In contrast, practitioners familiar with basic regression can be expected to relatively easily grasp the concepts of distributional regression — after all, in most of the corresponding models the main difference to linear regression is simply that we model not only the mean but also the variance. In any case, the resulting cost-optimal q_t can directly be “read” from the forecast distribution as implied by the fitted GAMLSS. To illustrate this point, Fig. 9 presents the GAMLSS with normally distributed response fitted to the data collected on the SKU grapes, here showing the associated forecast distributions for each weekday in the demand period 24-07-2017 to 29-07-2017 for FC 1. The model captures the fact that demand peaks at the beginning of a week, but also accounts for the associated increased variance in demand. The latter information would not readily be available when using quantile regression methods, as these target only at specific points of the forecast distribution, namely the quantiles of interest. Thus, the GAMLSS approach provides a more comprehensive picture of the demand to be expected.

6. Conclusion and future research

In this paper, we present a case study for distributional demand forecasting in e-grocery. For highly perishable SKUs with complex demand patterns, we find that models from the GAMLSS class tend to outperform the benchmark models, which is due to their increased flexibility in accommodating complex demand patterns. While our work was motivated by the new types of data found in e-grocery retailing, in particular uncensored demand and known demand at replenishment order time, all methods considered are, in principle, also applicable to traditional retail practice. Testing the GAMLSS approach for more SKUs, from diverse retailers, is generally desirable to analyse the benefits in retail practice, considering also the issue of robustness and scalability, as discussed in Fildes et al. (2018).

While the performance of the GAMLSS class in our case study is promising, we also found that no individual model consistently outperformed the other candidate models. This is clearly due to the diverse demand patterns found across SKUs, e.g. for SKUs with limited or without heteroscedasticity. Overall, there are very many possible combinations of the mean demand, its standard deviation and the level of asymmetry in the costs that together determine the total costs for a given demand period. Our empirical results, whereby no model class outperforms other model classes across all demand attributes, is in line with the no-free-lunch theorem (Wolpert & Macready, 1997), which establishes that an improved performance of any algorithm in a given class of problems is offset by a decreased performance in another class of problems. For data sets showing a high diversity of demand patterns as in our empirical case, future research could explore methods for automated model selection and/or automated model combination (also known as model averaging).

It may also be worthwhile to relax the assumption that the customers' purchases of different SKUs are independent decision problems. Specifically, we estimated the expected-cost optimal q separately for each SKU at a given service level. However, if SKUs are subject to substitution, or are complementary SKUs, this approach may not lead to the estimated cost-optimal stock levels. A simple way to tackle this situation would be to artificially increase the service level targets. Alternatively, joint distributions can be estimated to predict individual customer baskets.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank two anonymous referees for their insightful comments, which were tremendously useful.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2019.11.029.

References

Agrawal, N., & Smith, S. A. (1996). Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics, 43*(6), 839–861. doi:10.1002/(SICI)1520-6750(199609)43:6<839::AID-NAV4>3.0.CO;2-5.

Anderson, E. T., Fitzsimons, G. J., & Simester, D. (2006). Measuring and mitigating the costs of stockouts. *Management Science, 52*(11), 1751–1763. doi:10.1287/mnsc.1060.0577.

Anupindi, R., Dada, M., & Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science, 17*(4), 406–423. doi:10.1287/mksc.17.4.406.

Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal inventory policy. *Econometrica, 19*(3), 250–272. doi:10.2307/1906813.

Baker, K. R., Magazine, M. J., & Nuttle, H. L. W. (1986). The effect of commonality on safety stock in a simple inventory model. *Management Science, 32*(8), 982–988. doi:10.1287/mnsc.32.8.982.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. doi:10.1023/A:1010933404324.

Breiman, L., Cutler, A., Liaw, A., & Wiener, M. (2018). Package 'randomForest',. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Burgin, T. A. (1975). The gamma distribution and inventory control. *Operational Research Quarterly, 26*(3), 507–525. doi:10.2307/3008211.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research, 184*(3), 1140–1154. doi:10.1016/j.ejor.2006.12.004.

Conrad, S. (1976). Sales data and the estimation of demand. *Operational Research Quarterly, 27*(1), 123–127. doi:10.2307/3009217.

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting, 29*(3), 510–522. doi:10.1016/j.ijforecast.2012.09.002.

Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management, 18*(1), 69–88. doi:10.1287/msom.2015.0561.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*(1), 3–23. doi:10.1016/j.ijforecast.2008.11.010.

Fildes, R., Ma, S., & Kolassa, S. (2018). Retail forecasting: Research and practice. *Working Paper*. doi:10.13140/RG.2.2.17747.22565.

Godfrey, G. A., & Powell, W. B. (2001). An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science, 47*(8), 1029–1172. doi:10.1287/mnsc.47.8.1101.10231.

Hastie, T. J., & Tibshirani, R. J. (1986). Generalized Additive models. *Statistical Science, 1*(3), 297–318.

Haupt, H., Kagerer, K., & Steiner, W. J. (2014). Smooth quantilebased modeling of brand sales, price and promotional effects from retail scanner panels. *Journal of Applied Econometrics, 29*(6), 1007–1028. doi:10.1002/jae.2347.

Hohberg, M., Peter, P., & Kneib, T. (2018). Generalized additive models for location, scale and shape for program evaluation: A guide to practice. *Working Paper*. https://arxiv.org/pdf/1806.09386.pdf.

Kawamura, H. K., Nomoto, K. N., & Kuo, E. K. (2015). Inventory management based on demand forecasting using Ryokan's beer sales data. *Innovation and Supply Chain Management, 9*(4), 127–135. doi:10.14327/iscm.9.127.

Khouja, M. (1999). The single-period (news-vendor) problem: Literature review and suggestions for future research. *Omega, 27*(5), 537–553.

Klein, N., Kneib, T., Lang, S., & Sohn, A. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *Annals of Applied Statistics, 9*(2), 1024–1052. doi:10.1214/15-AOAS823.

Kneib, T. (2013). Beyond mean regression. *Statistical Modelling, 13*(4), 275–303. doi:10.1177/1471082X13494159.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives, 15*(4), 143–156. doi:10.1257/jep.15.4.143.

Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting, 32*(3), 788–803. doi:10.1016/j.ijforecast.2015.12.004.

Lahouar, A., & Ben Hadj Slama, J. (2015). Day-ahead load forecast using random forest and expert input selection. *Energy Conversion and Management, 103*, 1040–1051. doi:10.1016/j.enconman.2015.07.041.

Lang, S., Steiner, W. J., Weber, A., & Wechselberger, P. (2015). Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *European Journal of Operational Research, 246*(1), 232–241. doi:10.1016/j.ejor.2015.02.047.

Lau, H.-S., & Lau, A. H.-L. (1996). Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research, 92*(2), 254–265. doi:10.1016/0377-2217(95)00134-4.

Lu, C. J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing, 128*, 491–499. doi:10.1016/j.neucom.2013.08.012.

Ludwig, N., Feuerriegel, S., & Neumann, D. (2015). Putting Big Data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. *Journal of Decision Systems, 23*(1), 19–63. doi:10.1080/12460125.2015.994290.

Maciejowska, K., Nowotarski, J., & Weron, R. (2016). Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *International Journal of Forecasting, 32*(3), 957–965. doi:10.1016/j.ijforecast.2014.12.004.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science, 29*(9), 987–996. doi:10.1287/mnsc.29.9.987.

Matteo, A., & Wood, S. N. (2019). Package 'qgam', Retrieved from https://cran.r-project.org/web/packages/qgam/qgam.pdf.

Mayr, A., Fenske, N., Hofner, B., Kneib, T., & Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data-a flexible approach based on boosting. *Applied Statistics, 61*(3), 403–427. doi:10.1111/j.1467-9876.2011.01033.x.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research, 7*, 983–999. Retrieved from http://www.jmlr.org/papers/volume7/meinshausen06a/meinshausen06a.pdf.

Nahmias, S. (1994). Demand estimation in lost sales inventory systems. *Naval Research Logistics, 41*(6), 739–757. doi:10.1002/1520-6750(199410)41:6<739::AID-NAV3220410605>3.0.CO;2-A.

Nelder, A. J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, 135*(3), 370–384. doi:10.2307/2344614.

Ocado Group (2015). Delivering the best platform for online grocery. *Technical Report*. Retrieved from http://www.ocadogroup.com/~/media/Files/O/Ocado-Group/documents/fy15-annual-report-2015-v2.pdf

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research, 213*(2), 361–374. doi:10.1016/j.ejor.2010.11.024.

Ramaekers, K., & Janssens, G. K. (2008). On the choice of a demand distribution for inventory management models. *European Journal of Industrial Engineering, 2*(4), 479–491. doi:10.1504/EJIE.2008.018441.

Rigby, B., & Stasinopoulos, M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C Applied Statistics, 54*(3), 507–554. doi:10.1111/j.1467-9876.2005.00510.x.

Sachs, A. L., & Minner, S. (2014). The data-driven newsvendor with censored demand observations. *International Journal of Production Economics, 149*, 28–36. doi:10.1016/j.ijpe.2013.04.039.

Scarf, H., Arrow, K. J., & Karlin, S. (1959). A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, & H. Scarf (Eds.), *Studies in the mathematical theory of inventory and production* (pp. 201–209). California: Stanford University Press.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. doi:10.1214/10-STS330.

Song, J. S. (1994). The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Science, 40*(5), 603–613. doi:10.1287/mnsc.40.5.603.

Steiner, W. J., Brezger, A., & Belitz, C. (2007). Flexible estimation of price response functions using retail scanner data. *Journal of Retailing and Consumer Services, 14*(6), 383–393. doi:10.1016/j.jretconser.2007.02.008.

Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research, 178*(1), 154–167. doi:10.1016/j.ejor.2006.02.006.

Trusov, M., Bodapati, A. V., & Cooper, L. G. (2006). Retailer promotion planning: Improving forecast accuracy and interpretability. *Journal of Interactive Marketing, 20*(3–4), 71–81. doi:10.1002/dir.20068.

Weber, A., Steiner, W. J., & Lang, S. (2017). A comparison of semiparametric and heterogeneous store sales models for optimal category pricing. *OR Spectrum*, (2), 403–445. doi:10.1007/s00291-016-0459-6.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82. doi:10.1109/4235.585893.

Zipkin, P. H. (2000). *Foundations of inventory management* (1st). Boston: McGraw-Hill.