

**Universität Bielefeld**

**Faculty of Business Administration and Economics**

**Working Papers in Economics and Management**

No. 09-2018  
December 2018

## **Distributional regression for demand forecasting in e-grocery**

**Matthias Ulrich**

**Hermann Jahnke**

**Roland Langrock**

**Robert Pesch**

**Robin Senge**

Bielefeld University  
P.O. Box 10 01 31  
33501 Bielefeld – Germany

ISSN 2196–2723

ABCD [www.wiwi.uni-bielefeld.de](http://www.wiwi.uni-bielefeld.de)

# Distributional regression for demand forecasting in e-grocery

Matthias Ulrich <sup>1</sup>, Hermann Jahnke <sup>1</sup>, Roland Langrock <sup>1</sup>,  
Robert Pesch <sup>2</sup>, Robin Senge <sup>2</sup>

## Abstract

E-grocery offers customers an alternative to traditional brick-and-mortar grocery retailing. Customers select e-grocery for convenience, making use of the home delivery at a selected time slot. In contrast to brick-and-mortar retailing, in e-grocery on-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, compared to brick-and-mortar retailing, on-stock availability of SKUs has a 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 linear regression, quantile regression, and some popular methods from machine learning. The models are evaluated in a case study, where we compare their out-of-sample predictive performance with regard to the service level selected by the e-grocery retailer considered.

**Keywords:** Forecasting, Inventory, E-commerce, Retailing

---

<sup>1</sup>Department of Business Administration and Economics, Bielefeld University  
Universitätsstraße 25, 33615 Bielefeld, Germany. Email: matthias.ulrich@uni-bielefeld.de  
<sup>2</sup>inovex GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, Germany

# 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 underage costs due to missed sales, whereas excess inventory generates short-term overage costs due to spoilage and operational inefficiencies. However, due to high opportunity costs in retail practice, a service level selected exclusively based on short-term costs may not be profitable. 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 et al., 2006).

Within the fast-expanding electronic grocery (e-grocery) market, which provides an alternative to traditional brick-and-mortar 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 brick-and-mortar-store is required. However, if customers are unsatisfied with on-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 brick-and-mortar store to buy products affected by stock-outs. For perishable SKUs such as fruits, vegetables and meat, customers are particularly restricted with respect to potential substitutes, which further increases the risk of an order cancellation. On-stock availability can hence be expected to be of crucial importance regarding the customers' order decision. As a consequence, e-grocery retailers operate with very high strategic service level targets, e.g. 97–99% in the case of the retailer considered 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).

The business problem under consideration in this paper requires customer demand forecasts for each stock keeping unit (SKU) in each local 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 underlying demand, rendering demand forecasting very challenging (Fildes et al.,

2018). Figure 1 illustrates a small subset of the data considered in our case study, namely customer demand for the SKU grapes in the months April to July 2017. 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 brick-and-mortar retailing. Businesses typically make use of the grocery delivery on Mondays to supply their employees or guests with fruits, coffee and correlated SKUs such as milk.

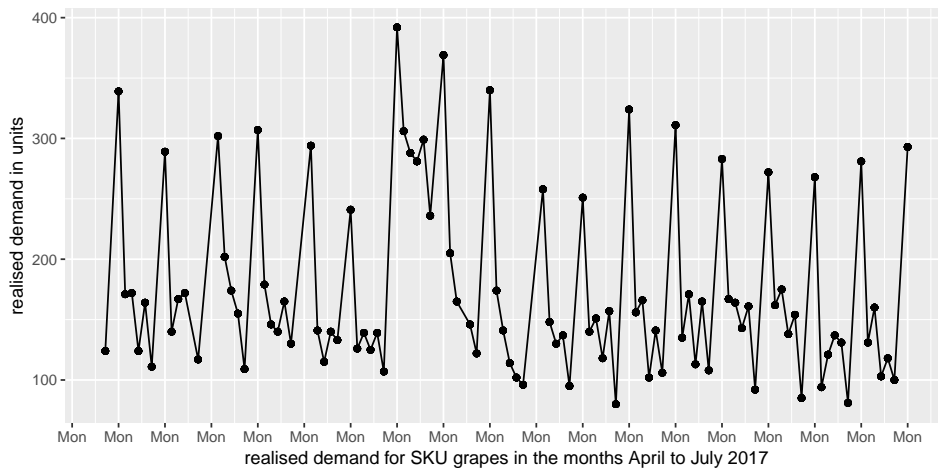


Figure 1: Realised demand for the SKU grapes in units (package of 500 gr), with highly variable demand but recurring peaks on Monday.

In traditional brick-and-mortar 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 one SKU has an impact also on demand data for correlated 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 SKUs affected by the correlation effect will therefore be potentially too low, whereas predicted future demand of selected substitutes will potentially be too high (Anupindi et al., 1998). In contrast

to traditional retailing, the online customer order process of the e-grocery retailer considered in our case study 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.

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 SKUs in grocery retailing, most results reported in the literature do indeed identify potential improvements by using nonlinear models (Ali et al., 2009 and Lang et al., 2015). In fact, 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. Quantile regression, where specific quantiles of the response variable (here: demand) are linked to covariates, constitutes a much more obvious alternative given the high service level targets (Manning et al., 1995 and 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 et al., 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 to consider various aspects of the response distribution (e.g. Mayr et al., 2012 and Klein et al., 2015). In contrast to quantile regression methods, which are nonparametric, distributional regression methods are parametric. Sachs & 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 if there is little censoring in the demand.

Considering data from a leading German omni-channel retailer, here we aim to exploit the associated new types of data in e-grocery that are not available in traditional brick-and-mortar retailing, i.e. uncensored customer demand and partly known future demand data, to model the entire demand distribution. Specifically, 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, i.e. for 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 (parametric) linear regression and (nonparametric) quantile regression, evaluating the models by comparing their out-of-sample forecasting error at the selected service level with the corresponding cost values for underage and overage. Given the increasing popularity of machine learning methods (see, e.g., Carbonneau et al., 2008, Ferreira et al., 2016), we also include parametric random forests (Breiman, 2001) and nonparametric quantile regression forests (Meinshausen, 2006) as benchmarks. For the SKUs considered in our case study, we find that models from the GAMLSS class tend to outperform the benchmark models, 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 significant 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.

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 that 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,

$$E[C_t(q_t)] = hE(q_t - D_t)^+ + bE(D_t - q_t)^+, \quad (1)$$

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 period  $t$ .

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

$$q_t^* = \underset{q_t}{\operatorname{argmin}} E[C_t(q_t)] = F_t^{-1}(b/(b+h)), \quad (2)$$

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 (2) 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 predict demand  $q_t$  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 et al. (2011) for extensive literature reviews with different applications for individual problem characteristics.

Early papers from this research area assume that the demand distribution is completely known (Arrow et al., 1951). Scarf et al. (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 expected-cost optimal order quantity  $q$  focus on the estimation of the mean by a point prediction and derive the standard deviation from historic demand variations. Based on the expected demand pattern, different distributional assumptions are 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 et al., 1986).

Linear regression is the most established approach for modelling directed relationships, and as such is also a popular tool for demand point forecasting (see, e.g., Makridakis et al., 1998, and Weisberg, 2005). However, linear regression involves several restrictive assumptions, namely linearity of the predictor, homoscedasticity and, when forecast distributions are of interest, also normality of the response. For at least 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.

For point forecasts of the mean, 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).

Quantile regression is a natural nonparametric alternative which focusses on forecasting selected quantiles rather than the mean (Koenker & Hallock, 2001, Maciejowska et al., 2016). In particular, within quantile regression we avoid making any distributional assumption for the response. However, a potential problem with using quantile regression methods for e-grocery demand forecasting is that we often see extreme outliers in the realised demand data to which quantile methods may react more strongly than distributional regression models. These outliers are due to the high proportion of business-to-business transactions as discussed above. Meinshausen (2006) proposed quantile regression forests as a competitive alternative to quantile regressions.

### **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 Figure 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.

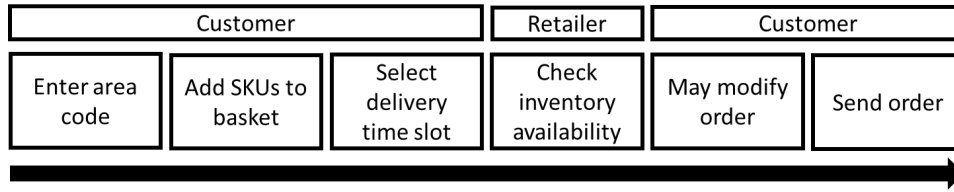


Figure 2: Customer order process in e-grocery.

The replenishment and fulfilment process of the e-grocery retailer is shown in Figure 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.

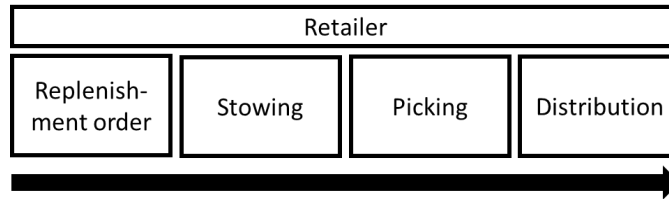


Figure 3: Replenishment and fulfilment process in e-grocery.

## 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, Figure 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 one FC. We find, *inter alia*, the following patterns:

1. *Nonlinearity*

Figure 4(a) shows the relationship between known customer demand at the

replenishment decision time and realised demand.<sup>1</sup> 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, i.e. above 100 units. This indicates that the functional relationship between realised demand and known demand is nonlinear.

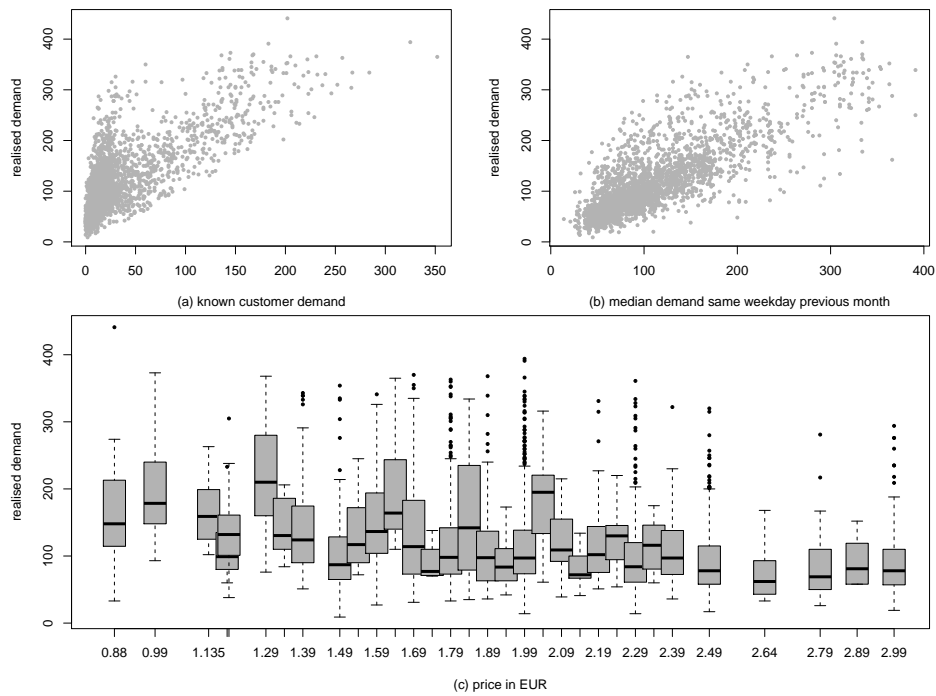


Figure 4: The relationship between selected explanatory variables (features) and the response variable, realised demand, for the SKU grapes within the months September 2015 to August 2017 for one FC.

<sup>1</sup>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.

## 2. *Heteroscedasticity*

Figure 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. The process of feature engineering is detailed in Section 3.2.

## 3. *Skewness*

Figure 4(c) shows positive skewness in the distribution of the realised demand, with the degree of skewness varying across different prices. The upper whisker and the 0.75 quantile are farther from the median than the 0.25 quantile and the lower whisker. This asymmetry in the distributional shape increases with increasing price.

# 3 Distributional regression

## 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 model is an extension of Generalised Linear Models (Nelder & Wedderburn, 1972) and Generalised Additive Models (Hastie & Tibshirani, 1986).

In GAMLSS, a parameter vector  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$  — 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  $y_i, i = 1, 2, \dots, n$ , are independent, each with an associated parameter vector  $\boldsymbol{\theta}^i = (\theta_{1i}, \theta_{2i}, \dots, \theta_{pi})$  and probability (density) function  $f(y_i|\boldsymbol{\theta}^i)$ . In our 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 & 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,

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \gamma_{jk}, \quad (3)$$

where  $g_k$  is applied componentwise,  $\boldsymbol{\theta}_k$  and  $\boldsymbol{\eta}_k$  are vectors of length  $n$ ,  $\mathbf{X}_k$  is a known design matrix of dimension  $n \times J'_k$ ,  $\boldsymbol{\beta}_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J'_k})^T$  is a parameter vector of length  $J'_k$ ,  $\mathbf{Z}_{jk}$  is a fixed known  $n \times q_{jk}$  design matrix and  $\gamma_{jk}$  is a  $q_{jk}$ -dimensional random variable. Thus, for  $k = 1, \dots, p$ , in the most general case the linear predictor  $\boldsymbol{\eta}_k$  includes the parametric component  $\mathbf{X}_k \boldsymbol{\beta}_k$  and in addition additive components  $\mathbf{Z}_{jk} \gamma_{jk}$ . Rigby & Stasinopoulos (2005) call model (3) the GAMLSS. The first two distributional parameters  $\theta_1$  and  $\theta_2$  are typically characterised as location and scale parameter, denoted by  $\mu$  and  $\sigma$ , respectively, whereas the remaining parameters are generally characterised as shape parameters.

For  $\mathbf{Z}_{jk} = \mathbf{I}_n$ , where  $\mathbf{I}_n$  is an  $n \times n$  identity matrix, and  $\gamma_{jk} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$  for all combinations of  $j$  and  $k$ , the GAMLSS model formulation (3) is semi-parametric given by

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \quad (4)$$

where the  $\mathbf{x}_{jk}$  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  $\mathbf{x}_{jk}$ . The explanatory vectors  $\mathbf{x}_{jk}$  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.

GAMLSS allows fitting a variety of different continuous and discrete distributions. An overview is given by Rigby & 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 literature. As a consequence, these are the distributions that we consider in our case study (with the corresponding models labeled

as GAMLSS\_NO, GAMLSS\_GA, GAMLSS\_PO and GAMLSS\_NB, respectively). GAMLSS offers two negative binomial distributions that differ in the definition of the variance. Type I defines the variance by  $(\mu + \sigma\mu^2)$  and type II by  $(\mu + \sigma\mu)$ , given a mean  $\mu$  and a standard deviation  $\sigma$  (Rigby & Stasinopoulos, 2008). 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` by Rigby & Stasinopoulos (2008).

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, as shown already in Figure 4. More specifically, the model that was fitted is specified as follows:

$$\begin{aligned} y_i &\sim N(\mu_i, \sigma_i^2) \\ \mu_i &= h_1(x_i) \\ \log(\sigma_i) &= h_2(x_i), \end{aligned}$$

where  $y_i$  is the  $i$ -th realised demand and  $x_i$  is the median demand of the same weekday in the previous month. 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 Equation (3).

The model displayed in Figure 5, while simplistic, adequately addresses the heteroscedasticity shown in Figure 4 by estimating the variance  $\sigma_i^2$  to increase for increased values of the covariate. 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.

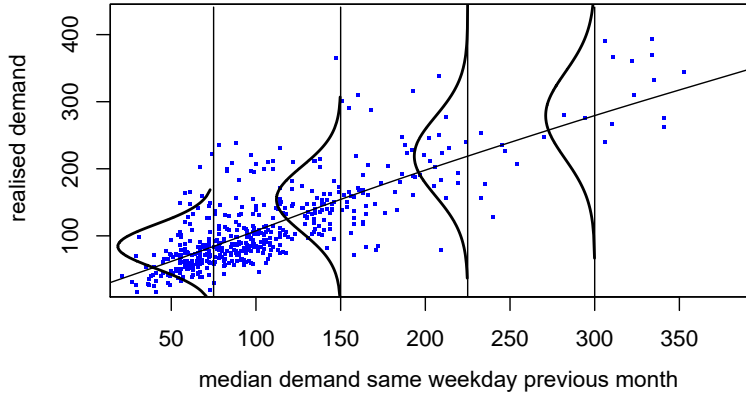


Figure 5: Realised demand as a function of median demand of the same weekday of the previous month, together with the fitted GAMLSS, including the distribution implied for the residuals at selected values of the covariate.

### 3.2 Feature engineering

For all models, the demand distribution  $F_t$  is estimated using features (explanatory variables). Feature engineering describes the process of generating suitable features from data. Both the general pattern of the demand distribution as well as any time series effects are taken into account by considering historic demand quantiles (5%, 50% and 95%), then building corresponding features using data from a) the previous quarter, b) the previous month, and c) the previous two weeks (Lu, 2014 and Kawamura et al., 2015). After feature engineering and feature selection the data set contains the following 12 features, including also price (Andreyeva et al., 2010) and known demand as extracted directly from the raw data.

1. Demand of the same weekday previous week in units
2. Demand of the same weekday second previous week in units
3. Median demand of the same weekday previous month in units
4. Median demand of the same weekday previous quarter in units
5. Median demand previous month in units

6. Median demand previous quarter in units
7. 95%-demand-quantile previous month in units
8. 95%-demand-quantile previous quarter in units
9. 5%-demand-quantile previous month in units
10. 5%-demand-quantile previous quarter in units
11. Price
12. Known customer demand in units

These features were selected based on previous analyses in the literature as well as business experience made within the retailer considered in our case study, and aim at capturing, *inter alia*, time series effects and overall variation in demand. The choices made are thus somewhat arbitrary, which however seems necessary given the lack of clear guidelines on feature engineering in the context of e-grocery demand forecasting. In any case, for our real-data analysis — the purpose of which is to compare methods for demand forecasting — the selection of features is not of primary importance. Feature selection using component-wise gradient boosting, as described in Hofner et al. (2016), did not improve the out-of-sample forecast accuracy in our case study, so that we eventually trained all models using the complete feature set.

### 3.3 Benchmark models

Based on our literature review in Section 2.2, we consider linear regression (LM), quantile regression (QR), random forests (RF), and quantile regression forests (QRF) as performance benchmarks for the GAMLSS approach proposed here.

Linear regression models estimate the model parameters by minimising the sum of squared residuals. Based on a fitted 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. Quantile regression methods directly predict any *conditional quantile* of the response distribution (Koenker & Hallock, 2001). 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 for the residuals, where for two-parameter distributions we simply use a constant standard deviation as estimated from one year of training data. For the random forest-based predictions we consider the same distributions as for GAMLSS (i.e. normal, gamma, Poisson, and negative binomial). Before applying the gamma distribution, we transformed the estimated mean and standard deviation to derive the required scale and shape parameter. Finally, quantile regression forests constitute a nonparametric alternative to random forests that allows the estimation of any *conditional quantile* (Meinshausen, 2006). The R packages we use are *rq* (Koenker, 2012) for quantile regression, *quantregForest* (Meinshausen, 2006) for the quantile regression forest, *caret* (Kuhn, 2008) for random forests and *gamlss* (Rigby & Stasinopoulos, 2008) for GAMLSS.

Thus, we apply the following benchmark models: Linear regression (LM); random forests with normal (RF\_NO); gamma (RF\_GA); Poisson (RF\_PO) and negative binomial distribution (RF\_NB); quantile regression (QR) and quantile regression forests (QRF).

## **4 Model training, demand forecasting and performance evaluation**

### **4.1 Training data**

Our data set covers demand periods from September 2015 to August 2017 and six different e-grocery FC. 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.

We train each model based on twelve months of data to subsequently forecast each 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 a corresponding sliding-window approach, we always train on the most recent twelve months of data. In particular, the length of twelve months of the training data set allows the incorporation of seasonal effects. 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  $> 300K$  predictions of  $q_t$  when considering 11 models and 99 possible service levels.

## 4.2 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 each model, and additionally the realised demand, for the 23 demand periods in April 2017. We visualise the different model classes by different symbols and the different distributional assumptions by different colours. 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 generated by a single FC, four models, namely GAMLSS\_PO, RF\_GA, RF\_NO and RF\_PO, yield an inventory level  $q_t$  that is below realised demand for at least two demand periods. Two demand periods with inventory levels below realised demand result in a service level of 91% (for this month). All other models yield inventory levels that in April 2017 never fall short of realised demand. Among these, considering overage costs for each excess unit of overage, we prefer inventory levels that are equal to or only slightly above realised demand. The inventory levels  $q_t$  obtained from GAMLSS\_GA, GAMLSS\_NBI, QR and RF\_NB show high positive deviations from realised demand. Thus, in this small subset of predictions, the

models GAMLSS\_NBII, GAMLSS\_NO, LM and QRF appear to be most promising with respect to minimising total costs for the SKU grapes.

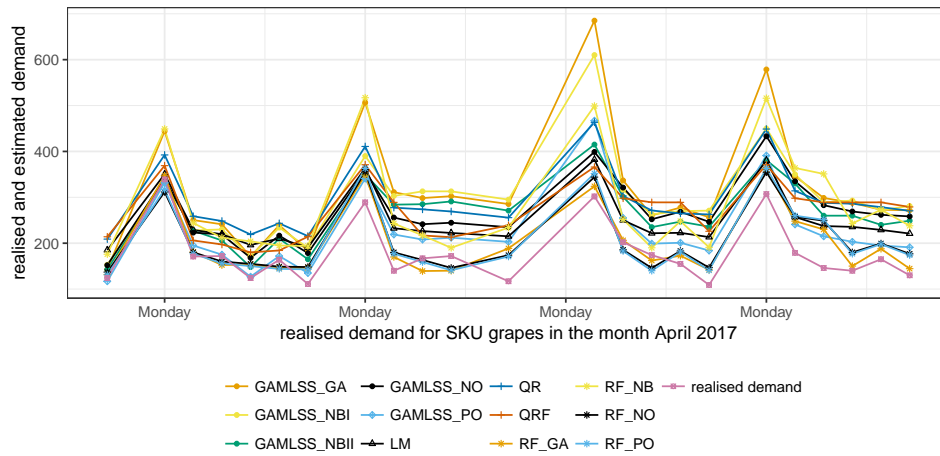


Figure 6: Predicted 0.97 demand quantiles as obtained with the different approaches and realised demand for the SKU grapes in April 2017.

To obtain a more comprehensive picture, Figure 7 illustrates the overall deviation between the service level target and the realised service level for each model and each service level target between 1% and 99%. If a model would meet each service level target exactly, then the corresponding line would follow the 45° line. The figure demonstrates that the models GAMLSS\_PO, RF\_GA, RF\_NO and RF\_PO lead to substantial deviations for any service level target distinct from 50%, i.e. whenever costs are asymmetric. More specifically, below the 50% service level target, the inventory levels as obtained by these models are too high, whereas they are too low for all service level targets above 50%.

In practice, only the service level target of the e-grocery retailer is pertinent for model evaluation. Figure 8 shows the percentage deviation between the realised service level and the 97% service level target. For this particular service level, we observe that only the the model QRF obtains a service level target above 97%. Overall, we identify that the nonparametric models, QR and QRF, are closer to the service level targets than the parametric models, which would seem to corroborate the results of Sachs & Minner (2014). The models GAMLSS\_PO, RF\_GA, RF\_NO

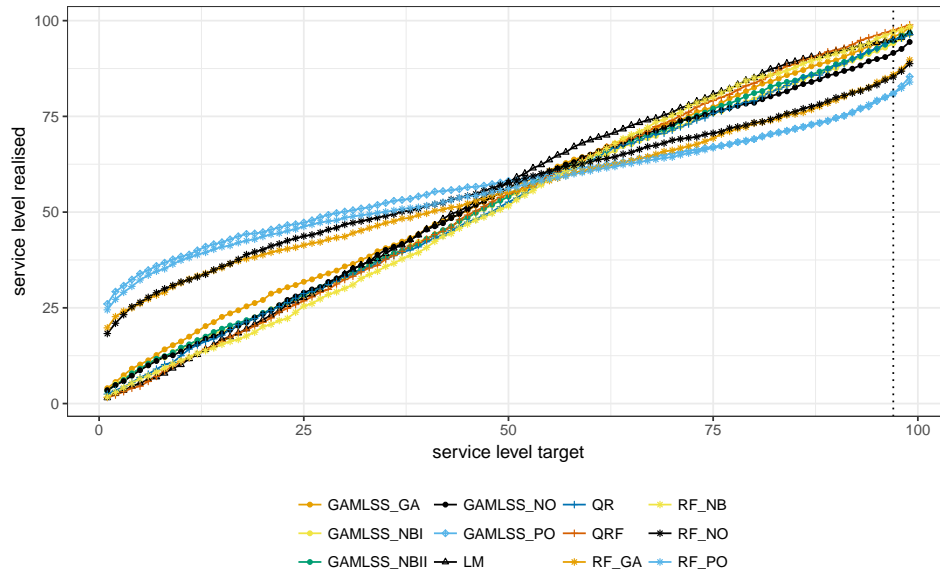


Figure 7: Service level target vs. realised service level for the SKU grapes in the validation period August 2016 to July 2017.

and RF\_PO show deviations of up to 17%. All other models show promising results in achieving the service level target with deviations below 6%, and hence are candidates for cost minimisation after considering asymmetric underage and overage costs.

### 4.3 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 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 underage and overage costs are asymmetric in our business problem — and we thus have to evaluate the two directions of error asymmetrically — we aim

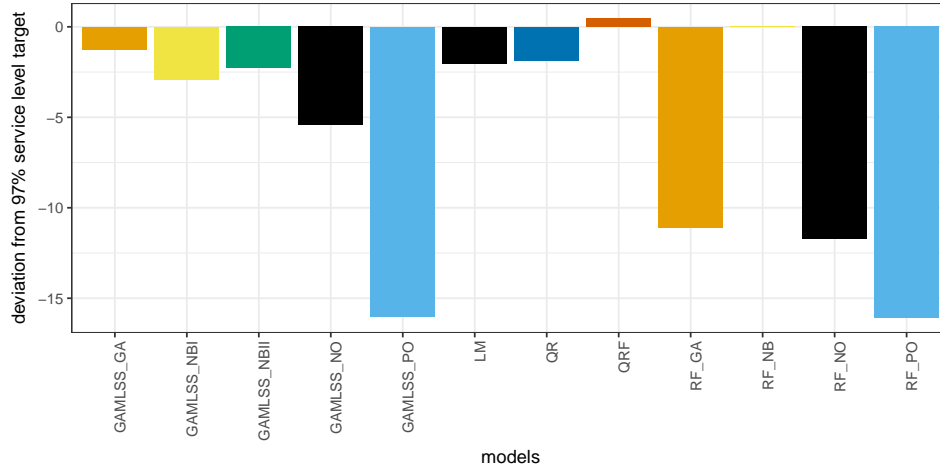


Figure 8: Model deviation from 97% service level target for the SKU grapes in the months August 2016 to July 2017.

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 overage cost parameter  $h$  is relatively straightforward given corresponding information from the retailer's accounting department, this is not the case for the underage 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 underage cost parameter  $b$  via the relation  $\alpha = b/(b+h)$  from the newsvendor problem, where  $\alpha$  is the service level. Given the empirical value of the service level and an estimation of the overage cost parameter  $h$  based on the retailer's margin and operational costs, we derive  $b$  as

$$b = \frac{h\alpha}{1-\alpha}.$$

Given the resulting cost values for  $b$  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 asymmetric cost values of  $b$  and  $h$  have a strong impact on the evaluation of the relative model performance. Assuming, as an example, overage costs of 1 EURO and a service level target of 97%, we obtain underage costs of 32,30 EUR for each unit of underage. 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

Figure 9 shows the percentage difference in total costs, according to (1), as obtained under the different approaches, in each case compared to the benchmark linear regression:

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

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

For all SKUs considered, we find that models from the GAMLSS class outperform the benchmark models, with different distributions across the different SKUs yielding the lowest out-of-sample costs. Compared to linear regression, we find potential cost reductions of up to 25%. 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. GAMLSS allows us to tailor the regression model to whatever pattern we find for a given SKU, such that, as expected, in most cases it leads to improved demand forecasts. An exception is the model GAMLSS\_PO, which in terms of its single-parameter response distribution is actually less flexible than a basic, possibly non-linear regression model — the implied mean-variance relation of the Poisson here drastically limits the model's ability to capture the empirical patterns. Overall, the results demonstrate the strengths of

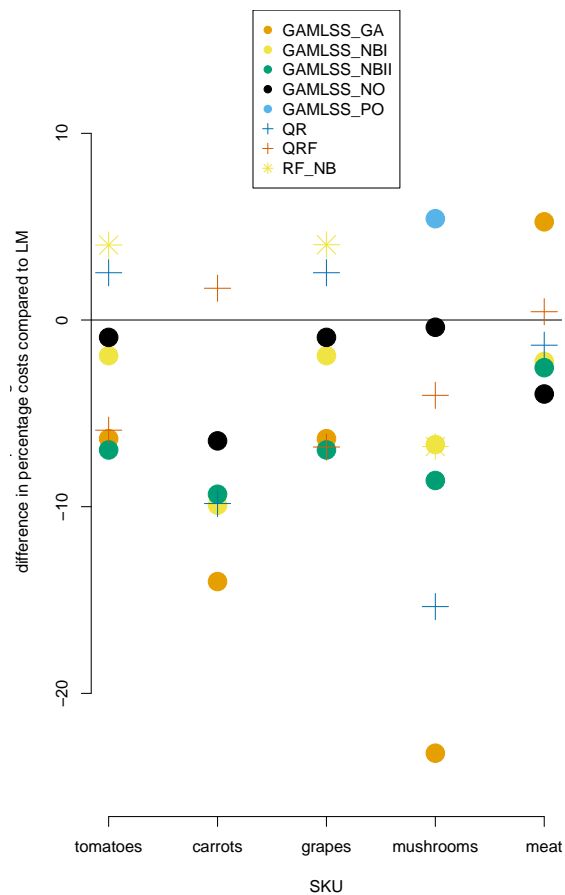


Figure 9: Relative total costs as obtained under the different approaches, compared to the benchmark linear regression, for five different SKUs.

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 (e.g. fast-moving or slow-moving SKUs with regular or irregular demand).

Competing model classes such as random forests also allow a flexible selection

of distributions. However, with our somewhat ad hoc approach to complement the point forecast by a distributional assumption for the residuals, 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. Random forests do indeed yield much higher costs across all distributions considered. In contrast to random forests, the quantile regression-based approaches performed very well in our case study. However, due to their sensitivity especially for extreme quantiles (e.g. 0.99), they still do bear the risk of occasionally producing demand forecasts that are way off. In addition, it can be disadvantageous 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.

In our view, interpretability in retailing practice is in fact an important benefit of the GAMLSS class. 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 adjustments in the case of exceptional circumstances (cf. Fildes et al., 2009, Davydenko & Fildes, 2013). In particular, machine learning techniques are often restricted in their possibilities of interpretation. 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, Figure 10 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. 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.



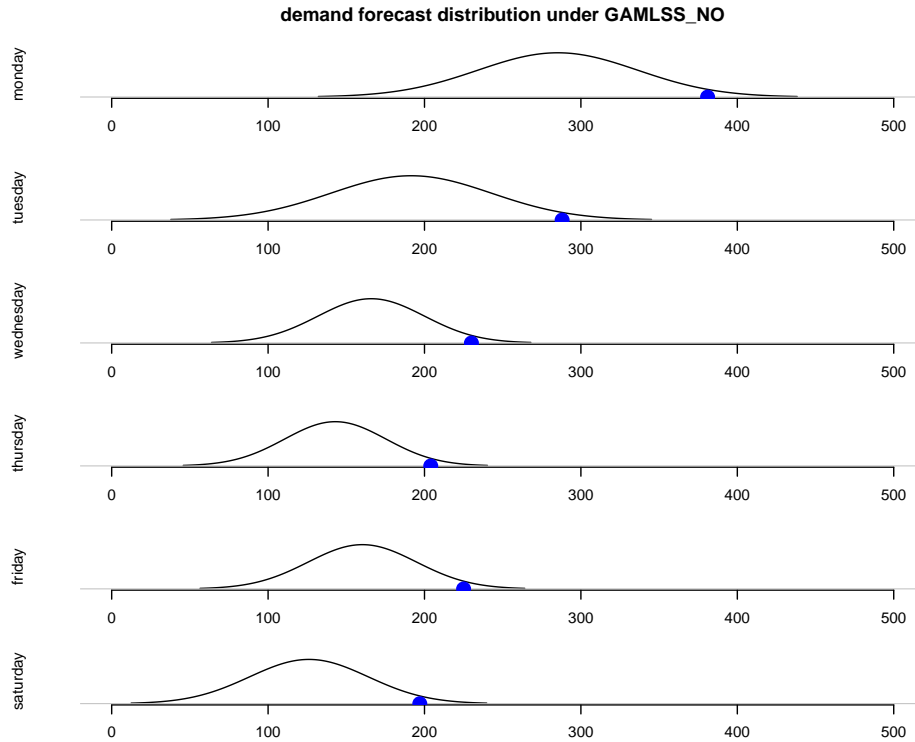


Figure 10: Demand forecast distribution for the SKU grapes under the model GAMLSS\_NO in the week 24-07-2017 to 29-07-2017; the 97% quantiles under each daily forecast distribution are highlighted in blue.

## 6 Conclusion and future research

In this paper, we present a case study for distributional demand forecasting in e-grocery. For highly perishable SKU 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 accommdating 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.

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 establish 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 should, therefore, focus on methods for automated model selection and/or automated model combination (also known as model averaging). The various candidate models for demand forecasting have different strengths and weaknesses in capturing particular demand patterns. Thus, we believe that there is considerable potential to reduce economic costs by developing algorithms that automate the model selection and possibly also a model averaging process.

An additional direction of research may relax the assumption of demand independence between SKUs. In this paper we estimated the expected-cost optimal  $q$  separately for each SKU at a selected service level. However, if SKUs are subject to significant substitution or correlation effects this approach may not lead to the estimated cost optimal stock levels. One simple way to tackle this situation may be to artificially increase the service level targets. Alternatively, the estimation of a joint distribution function could be considered, which however also requires the estimation of individual customer baskets.

## 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: [https://doi.org/10.1002/1520-6750\(199609\)43:6::AID-NAV3220430601>3.0.CO;2-Z](https://doi.org/10.1002/1520-6750(199609)43:6::AID-NAV3220430601>3.0.CO;2-Z)

- Ali, Ö. G., Sayin, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10), 12340–12348. doi: <https://doi.org/10.1016/j.eswa.2009.04.052>
- Anderson, E. T., Fitzsimons, G. J., & Simester, D. (2006). Measuring and mitigating the costs of stockouts. *Management Science*, 52(11), 1751–1763. doi: <https://doi.org/10.1287/mnsc.1060.0577>
- Andreyeva, T., Long, M. W., & Brownell, K. D. (2010). The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food. *American Journal of Public Health*, 100(2), 216–222. doi: <https://doi.org/10.2105/AJPH.2008.151415>
- 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: <https://doi.org/10.1287/mksc.17.4.406>
- Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal inventory policy. *Econometrica*, 19(3), 250–272. doi: <https://doi.org/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: <https://doi.org/10.1287/mnsc.32.8.982>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi: <https://doi.org/10.1023/A:1010933404324>
- Burgin, T. A. (1975). The gamma distribution and inventory control. *Operational Research Quarterly*, 26(3), 507–525. doi: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/10.1016/j.ijforecast.2008.11.010>

Fildes, R., Ma, S., & Kolassa, S. (2018). Retail forecasting: Research and practice. *Working Paper*. doi: <https://doi.org/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: <https://doi.org/10.1287/mnsc.47.8.1101.10231>

Hastie, T., & Tibshirani, R. J. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–318. doi: <https://doi.org/10.1214/ss/1177013609>

Hofner, B., Mayr, A., & Schmid, M. (2016). gamboostLSS: An R package for model building and variable selection in the gamlss framework. *Journal of Statistical Software*, 74(1), 1–31. doi: <https://doi.org/10.18637/jss.v074.i01>

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

10.1016/S0305-0483(99)00017-1

- 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: <https://doi.org/10.1214/15-AOAS823>
- Kneib, T. (2013). Beyond mean regression. *Statistical Modelling*, 13(4), 275–303. doi: <https://doi.org/10.1177/1471082X13494159>
- Koenker, R. (2012). Package ‘quantreg’. Retrieved from <https://cran.r-project.org/web/packages/quantreg/quantreg.pdf>
- Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156. doi: <https://doi.org/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: <https://doi.org/10.1016/j.ijforecast.2015.12.004>
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. doi: <https://doi.org/10.18637/jss.v028.i05>
- 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: <https://doi.org/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: <https://doi.org/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: [https://doi.org/10.1016/0377-2217\(95\)00134-4](https://doi.org/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: <https://doi.org/10.1016/j.neucom.2013.08.012>

- 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: <https://doi.org/10.1016/j.ijforecast.2014.12.004>
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications* (3rd ed.). John Wiley & Sons, Inc.
- Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, 29(9), 987–996. doi: <https://doi.org/10.1287/mnsc.29.9.987>
- Manning, W. G., Blumberg, L., & Moulton, L. H. (1995). The demand for alcohol: The differential response to price. *Journal of Health Economics*, 14(2), 123–148. doi: [https://doi.org/10.1016/0167-6296\(94\)00042-3](https://doi.org/10.1016/0167-6296(94)00042-3)
- 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: <https://doi.org/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: [https://doi.org/10.1002/1520-6750\(199410\)41:6<739::AID-NAV3220410605>3.0.CO;2-A](https://doi.org/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: <https://doi.org/10.2307/2344614>
- Ocado Group. (2015). Delivering the best platform for online grocery. (November). Retrieved from <http://www.ocadogroup.com/{~}/media/Files/0/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: <https://doi.org/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: <https://doi.org/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: <https://doi.org/10.1111/j.1467-9876.2005.00510.x>

Rigby, B., & Stasinopoulos, M. (2008). GAMLSS package: A flexible regression approach. Retrieved from <https://pdfs.semanticscholar.org/ed3d/f96aa477d6454d5737403a5aab751ccd9315.pdf>

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

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

Weisberg, S. (2005). *Applied Linear Regression* (3rd ed.). A John Wiley & Sons, INC., Publication.

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

Zipkin, P. H. (2000). *Foundations of Inventory Management* (1st ed.). McGraw-Hill, Boston.