International Journal of Production Research, 2018
Vol. 56, No. 12, 4168–4185, https://doi.org/10.1080/00207543.2018.1431413

Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions

Ville Sillanpää* and Juuso Liesiö

Department of Information and Service Economy, Aalto University, Helsinki, Finland

(*Received 7 April 2017; accepted 15 January 2018*)

In retail, distribution centres can forecast the stores' future replenishment orders by computing planned orders for each stock-keeping-unit. Planned orders are obtained by simulating the future replenishment ordering of each stock-keeping-unit based on information about the delivery schedules, the inventory levels, the order policies and the point-estimate forecasts of consumer demand. Point-estimate forecasts are commonly used because automated store ordering systems do not provide information on the demand distribution. However, it is not clear how accurate the resulting planned orders are in the case of products with low and intermittent demand, which make up large parts of the assortment in retail. This paper examines the added value of modelling consumer demand with distributions, when computing the planned orders of products with low and intermittent demand. We use real sales data to estimate two versions of a planned order model: One that uses point-estimates and another that uses distributions to model the consumer demand. We compare the forecasting accuracies of the two models and apply them to two example applications. Our results show that using distributions instead of point-estimates results in a significant improvement in the accuracy of replenishment order forecasts and offers potential for substantial cost savings.

Keywords: forecasting; replenishment orders; planned orders; retail supply chain; low and intermittent demand

1. Introduction

During the past 20–30 years many retail chains have adopted a mode of operation in which large distribution centres (DCs) supply the stores with majority of their assortment (Fernie, Pfab, and Marchant 2000; Fernie, Sparks, and McKinnon 2010; Kuhn and Sternbeck 2013). In recent years, automated store ordering systems have also become common in retail chains (see, e.g. van Donselaar et al. 2006; van Donselaar et al. 2010; Kuhn and Sternbeck 2013; Ehrenthal, Honhon, and Woensel 2014). In these systems, replenishment ordering decisions are taken automatically based on stock-keeping-unit (SKU)-specific inventory control policies and demand forecasts. If the stores share information on historical consumer sales, inventory levels and ordering policies with the DC, the DC can use material requirements planning models (see, e.g. Orlicky 1975) to forecast the stores' replenishment orders. Forecasts produced with material requirements planning models are often referred to as *planned orders* (Jonsson and Mattsson 2013), and they are computed by simulating the SKUs' future replenishment orders based on information about the inventory levels, the order policies, the delivery schedules and the point-estimate forecasts of consumer demand. Point-estimate forecasts are used in planned orders, because retail's automated store ordering systems predominantly use point-estimates to model consumer demand (van Donselaar et al. 2006, 2010). Byrne and Heavey (2006) report that notable cost savings can be achieved using planned orders, while Jonsson and Mattsson (2013) conclude that planned order information is valuable to the supplier when the demand is non-stationary.

However, it is unclear how accurate the planned orders are in forecasting the replenishment orders, when the consumer demand is low and intermittent. In their study on planned orders, Jonsson and Mattsson (2013) note that their results are not necessarily generalisable to products with low and intermittent demand. Moreover, several authors have shown that in many applications point-estimate forecasts are not suitable for modelling low and intermittent demand (e.g. Snyder, Ord, and Beaumont 2012; Chapados 2014). Specifically, such demand processes exhibit large numbers of zero realisations and large variance to mean ratios (Agrawal and Smith 1996), and it is unclear how these properties affect the accuracy of planned orders. This is particularly important in retail, where products with low and intermittent demand often make up a large part of the assortment (see, e.g. Småros et al. 2004; Ketzenberg and Ferguson 2008). This observation also applies to our case company, a European grocery retailer, who reports that 49% of its SKUs have mean daily demand of less than 0.2 units.

In this paper, we investigate how accurately planned orders predict the replenishment orders of SKUs with low and intermittent demand. Furthermore, we investigate how this accuracy could be improved. Towards this end, we examine two versions of a planned order model designed for forecasting the replenishment orders of retail outlets: in the first version, which

*Corresponding author. Email: ville.sillanpaa@aalto.fi

represents the current practice of a European grocery retailer, the future consumer demand is modelled with point-estimate forecasts. The second version extends the first one by allowing the consumer demand forecasts to be distributions.

We compare the order forecasts produced with the two models using real sales data from a European grocery retailer. Our results show that using distribution forecasts offers a significant improvement in forecasting accuracy. To estimate the monetary value of the improved forecasts, we apply the models to a workforce planning problem and an inventory management problem. Results of these tests indicate that the improved forecast accuracy can translate to substantially lower costs in practical applications at the DC.

The distribution based planned order model of our paper also offers a practical tool to support decision-making in retail supply chains. Sternbeck and Kuhn (2014) and Holzapfel et al. (2015) have shown that retail chains can achieve significant cost savings through better design of delivery schedules. However, changing the delivery schedules requires methods for analysing how the changes affect the volumes of replenishment orders the stores place to the DC. Our results show that the distribution-based planned order model proposed in this paper is accurate enough for conducting such analyses.

The rest of this paper is organised as follows. Section 2 reviews existing literature on order forecasting and forecasting low and intermittent demand. Section 3 develops the point-estimate and the distribution based planned order models. In Section 4, the performances of the two models are compared using real data, and in Section 5 the results of the comparison are discussed. In Section 6, conclusions are drawn and future research directions are suggested.

2. Literature review

2.1 *Earlier approaches to order forecasting*

Suppliers require forecasts of incoming orders to effectively plan their operations, which, for instance, include inventory control (Jonsson and Mattsson 2013) and order picking (De Koster, Le-Duc, and Roodbergen 2007). A key component of any forecast is information: not surprisingly plenty of supply chain research has focused on identifying the benefits and challenges of sharing information in supply chains (e.g. Waller, Johnson, and Davis 1999; Aviv 2002, 2007; Småros 2007; Sari 2008; Ketzenberg and Ferguson 2008; Cannella and Ciancimino 2010; Zhang and Cheung 2011; Giloni, Hurvich, and Seshadri 2014; Alftan et al. 2015).

A straightforward way to share information to suppliers is to share point-of-sales (POS) data, i.e. SKU-specific sales data. Forecast models that use POS data are based on the idea that consumer sales is a proxy for the store's ordering activity. This motivates using statistical time series models that predict the store's orders using the historical POS data. This approach order forecasting can be problematic, because the decision to order also depends on the store's current inventory level (Williams and Waller 2011). Williams et al. (2014) have addressed this issue by developing a model where POS data and supplier's order history are used together.

Forecast methods that use only POS data and historical orders are difficult to use when the delivery schedules are changed. Typically, the DC and the stores agree on delivery schedules i.e. fixed and regularly repeating windows for ordering each product from the DC (Kuhn and Sternbeck 2013; Sternbeck and Kuhn 2014). Because, the delivery schedules determine when the stores can order specific products, they have a major impact on the stores' ordering behaviour (Xu and Leung 2009). Models that rely only on POS data and order history assume that delivery schedules remain constant over time. Consequently, POS data and order history models cannot respond to changes in ordering that are caused by changes in delivery schedules. This is a major concern, because typically delivery schedules are changed in retail frequently. For example, our case company alters delivery schedules before Christmas or Easter because labour costs of warehouse picking are more expensive on holidays and because consumer demand increases during the holidays. Furthermore, Gaur and Fisher (2004), Esparcia-Alcazar et al. (2010), Sternbeck and Kuhn (2014) and Holzapfel et al. (2015) have observed that by optimising delivery schedules retail chains can reduce logistics costs. Thus, retailers have incentives to alter delivery schedules also outside the holiday seasons.

Using stores' planned orders as forecasts takes changes in the delivery schedules into account. Idea in planned orders is to forecast the stores' future replenishment ordering decisions by simulating them as functions of the consumer demand forecasts, the delivery schedules and the order policies. Because the delivery schedules are parameters in the estimation of planned orders, changes in delivery schedules are not a source of concern when planned orders are used for forecasting the stores' replenishment orders.

Planned orders are also useful when a retailer is analysing the effects large changes in consumer demand have on replenishment order volumes. In retail, SKU-level demand forecasts are often calculated for consumer sales units (van Donselaar et al. 2010; Trapero, Kourentzes, and Fildes 2015). If such a forecast is changed dramatically due to, for example, a chain-wide promotion, it is often not clear how the change affects the order volumes. This is because ordering decision is a function of many parameter values of which demand forecast is only one. Therefore, in the case where the demand forecasts are changed, planned orders are a way to map the changes in demand forecasts to the changes in the order volumes.

Several authors have done research on planned orders and reported that using them as forecasts benefits the supplier. Zhao, Xie, and Leung (2002) have tested planned orders under several different assumptions of demand, and concluded that forecasting with planned orders outperforms forecasting methods which incorporate only POS data or historical order data. Byrne and Heavey (2006) showed that using planned orders decreased supply chain costs of both the supplier and the retailer. Jonsson and Mattsson (2013) have examined the use of planned orders by demand type, and concluded that using them decreases supplier's inventory levels in the case of non-stationary demand, but increased them in the cases of stationary and seasonal demand.

Although planned orders are usually seen as a tool for suppliers, retail DCs can also compute planned orders of the retail stores' replenishment orders. This is possible, because the prevailing practice in retail is to automate replenishment ordering processes (Kuhn and Sternbeck 2013). When automated, retailers' replenishment ordering follows clearly defined policies, which makes forecasting with planned orders possible. In particular, the stores' replenishment ordering policies can be simulated based on current inventory levels, demand forecasts and the parameters of the replenishment policy.

While Zhao, Xie, and Leung (2002), Byrne and Heavey (2006) and Jonsson and Mattsson (2013) provide insights about the usefulness of planned orders, they do not discuss the performance of planned orders when demand is low and intermittent. Moreover, the aforementioned papers do not describe the computation of planned orders with a detail that would allow repeating the experiments. To our knowledge, only Montanari et al. (2015) has given a detailed specification for a planned order model. The model of Montanari et al. (2015) assumes that stores use an economic order quantity (EOQ)-based policy. Thus the model described in our paper is a Monte Carlo simulation based extension to the model by Montanari et al. (2015), since it produces order distribution forecasts for SKUs with a dynamic reorder point policy.

It is worth noting that DC's need forecasts of replenishment orders because optimising replenishment in retail chains is often challenging. Marklund (2002), Monthatipkul and Yenradee (2008) and Kiesmüller and Broekmeulen (2010) have developed optimisation models where the store's replenishment policy's parameters are used as decision variables to minimise the stores' and the DC's costs jointly. While these papers do not discuss forecasting with planned orders, they are related to the problem of forecasting replenishment orders, as the ultimate goal of forecasting is to improve operational efficiency. However, the optimal policies derived in these papers are not directly applicable to our case, because they all assume that the costs that incur from using the replenishment policy can be defined explicitly. While this is a common assumption in the inventory management literature, it is often difficult to define the costs in practice (Tiwari and Gavirneni 2007). In retail the customer responses to out-of-stocks depend on customer loyalty (Schary and Christopher 1979; Emmelhainz, Stock, and Emmelhainz 1991; Campo, Gijsbrechts, and Nisol 2000) and on situational factors such as the existence of substitutes (Campo, Gijsbrechts, and Nisol 2000; Zinn and Liu 2001; Zinn and Liu 2008). Hence, it not often clear what the costs of lost sales are.

Because costs of lost sales are difficult to estimate, retail chains tend to use heuristic order policies to replenish stores. As retail demand often exhibits seasonal variation between, e.g. weekdays (Ehrenthal, Honhon, and Woensel 2014), replenishing a retail SKU over time is essentially a dynamic lot-sizing problem (see, e.g. Wagner and Whitin 1958). Therefore, automated store ordering systems in retail often use periodic-review order policies with dynamic reorder points (van Donselaar et al. 2010; Kuhn and Sternbeck 2013), i.e. reorder points which change as functions of the consumer demand forecasts. In this paper, we do not consider the optimisation of the stores' or the DC's order policies. Instead, we develop a model which allows the DC to forecast the stores' replenishment orders as functions of the stores' heuristic ordering policies. Approaches for identifying optimal order policies under time-varying demand distributions have been developed by, e.g. Feng, Rao, and Raturi (2011).

2.2 *Forecasting low and intermittent consumer demand*

Computation of planned orders requires consumer demand forecasts as one input. Since the focus of this paper is on SKUs with low and intermittent demand, we review the literature on suitable forecast models. We omit papers which consider the problem of classifying products by demand type (see, e.g. Croston 1974; Williams 1984), since our focus is on examining how the type of consumer demand forecasts (point-estimate or distribution) affects the computation of planned orders.

Table 1 classifies some of the approaches developed for forecasting low and intermittent demand. In particular, plenty of research has focused on identifying the suitable parametric distribution for the demand as well as on estimating regression models under varying distribution assumptions. Syntetos, Babai, and Altay (2012), Syntetos, Lengu, and Babai (2013) tested several distribution assumptions with data on spare parts demand. The authors concluded that the Negative Binomial and the Gamma distributions offer the best performance in terms of inventory control, even though the Gamma distribution performed poorly in goodness of fit tests. Snyder, Ord, and Beaumont (2012) tested how well the Poisson, Negative Binomial and Hurdle-shifted Poisson describe the demand of spare parts. The authors concluded that the standard Poisson model performs poorly when demand is overdispersed. This is logical, since the Poisson model does not take overdispersion into account. Kolassa

Table 1. Approaches for forecasting low and intermittent demand.

Article	Forecasting approach	Demand distribution assumptions
Syntetos, Babai, and Altay (2012)	Assumed i.i.d demand over periods	Negative Binomial, Gamma, Normal, Poisson, Stuttering Poisson
Snyder, Ord, and Beaumont (2012)	Assumed i.i.d demand over periods, moving average	Negative Binomial, Poisson, Hurdle-Shifted Poisson
Kolassa (2016)	Croston, regression models, bootstrapping	Negative Binomial, Poisson, empirical
McKenzie (1985)	Autoregressive models	Poisson, Geometric, Negative Binomial
Al-Osh and Alzaid (1987)	Autoregressive models	Empirical
Zhang et al. (2014)	Autoregressive models	Poisson
Alwan and Weiß (2017)	Autoregressive models	Poisson, Negative Binomial
Dolgui, Pashkevich, and Pashkevich (2004)	Bayesian updating	Beta-Binomial
Dolgui, Pashkevich, and Pashkevich (2005)	Bayesian updating	Beta-Binomial
Dolgui and Pashkevich (2008)	Bayesian updating	Hierarchical Beta-Binomial
Croston (1972)	Croston	Normal (demand size), Geometric (arrivals)
Syntetos and Boylan (2001)	Croston	Normal (demand size), Geometric (arrivals)
Syntetos and Boylan (2005)	Croston	Normal (demand size), Geometric (arrivals)
Levén and Segerstedt (2004)	Croston	Erlang
Teunter, Syntetos, and Babai (2011)	Croston	Logarithmic (demand size), Poisson (arrivals)
Willemain, Smart, and Schwarz (2004)	Bootstrapping	Empirical
Syntetos, Babai, and Gardner (2015)	Croston, exponential smoothing, bootstrapping	Normal, Empirical
Nikolopoulos, Babai, and Bozos (2016)	Nearest neighbors	(Not applicable)
Nikolopoulos et al. (2011)	Temporal aggregation	(Not applicable)
Rostami-Tabar et al. (2013)	Temporal aggregation	Normal
Jin et al. (2015)	Temporal aggregation	(Not applicable)
Zhang, Zhang, and Liu (2016)	Aggregation over SKUs	(Not applicable)

(2016) tested a novel forecast accuracy metric by fitting various demand models with several distribution assumptions to a data-set of retail demand. Like Snyder, Ord, and Beaumont (2012), Kolassa (2016) concluded that the Poisson models fared poorly. Integer-valued autoregressive models have been developed for handling autocorrelated demand series (see McKenzie 1985; Al-Osh and Alzaid 1987), and also tested in applications (Zhang et al. 2014; Alwan and Weiß 2017).

Some methods for forecasting low and intermittent demand are build on the Bayesian approaches. For instance, Dolgui, Pashkevich, and Pashkevich (2004, 2005) and Dolgui and Pashkevich (2008) have modelled low and intermittent demand with a hierarchical Beta-Binomial model. This approach is suitable especially for situations, where the amount of data on individual SKUs is limited.

Methods for generating point-estimate forecasts for low and intermittent demand have received plenty of attention during the last 40 years (Syntetos, Boylan, and Disney 2009; Boylan and Syntetos 2010), with the Croston model being perhaps the most well-known approach (Croston 1972). In the Croston model, the mean demand and the time interval between successive demand realisations are forecasted separately using exponential smoothing. The Croston model has since been developed further: Syntetos and Boylan (2001) noted that the original Croston model is biased, and suggested alleviating this bias by adding a constant term to the equation used in updating the forecasts. Syntetos and Boylan (2005) developed an alternative bias correction, where the constant term depends on the exponential smoothing parameters. Levén and Segerstedt (2004) also propose a modification to the Croston forecasting procedure, and extend the procedure for modelling demand variance under the Erlang distribution. Teunter, Syntetos, and Babai (2011) suggest modelling the probability of a demand occurrence instead of the time interval between successive demand realisations. Kolassa (2016) has used a variant of the Croston model to produce forecasts based on the Poisson and the Negative Binomial distributions.

Nonparametric (e.g. bootstrap) methods have also been tested for forecasting low and intermittent demand (e.g. Willemain, Smart, and Schwarz 2004). However, Syntetos, Babai, and Gardner (2015) and Kolassa (2016) present evidence in favour of parametric methods over the more computationally intensive non-parametric methods.

Other recent developments for forecasting intermittent demand include models based on the nearest neighbour method, temporal aggregation and aggregation over the SKUs. Nikolopoulos, Babai, and Bozos (2016) develop a forecasting method based on k-nearest neighbours, but note that the model's forecasting performance deteriorates when the frequency of zero-

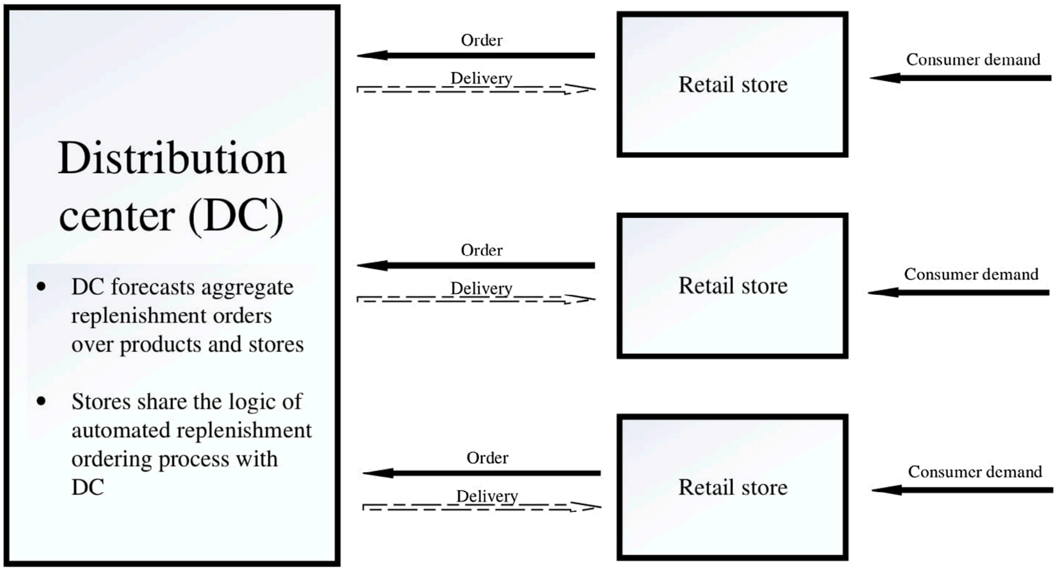

Figure 1. Visual summary of interaction between the DC and the retail stores. The stores place automated replenishment orders to respond to consumer demand. The stores share their forecasts and order policies with the DC, so that the DC can simulate the future replenishment ordering processes of stores for its own forecasting purposes.

observations increases. Temporal aggregation models have also recently attracted notable interest (see, e.g. Nikolopoulos et al. 2011; Rostami-Tabar et al. 2013; Jin et al. 2015). The key idea here is to aggregate the demand data from shorter time segments (e.g. a day) to longer time segments (e.g. a week), and then fit a statistical model to the aggregated time series. After forecasting on the aggregate level, the forecasts are distributed to the shorter time segments using some pre-specified distribution weights. Zhang, Zhang, and Liu (2016) have developed a model in which the demand series of SKUs are aggregated using the analytic hierarchy process (AHP). Generally, using aggregation models is slightly more complicated than using traditional ones, since the aggregation level and distribution weights have to be specified.

3. Planned order model for forecasting stores' replenishment orders

This section introduces a model for computing planned orders in a setting, where replenishment orders are placed by retail stores to a DC. Two versions of the model are considered: the first version uses point-estimate consumer demand forecasts and thereby serves as a benchmark representing the prevailing practice of many retail supply chains. Indeed, many existing automated store ordering systems provide a similar functionality for order forecasting. The second version extends the first one by allowing consumer demand forecasts to be distributions, which also necessitates some changes to the computation procedure.

Both versions of the model are essentially material requirements planning models (Orlicky 1975) where the future requirements of each store for each SKU are derived from the SKUs' consumer demand forecasts, delivery schedules and order policies. We acknowledge that similar models have been introduced earlier in the context of, e.g. manufacturing (see, e.g. Vollmann, Berry, and Whybark 1997; Koh, Saad, and Jones 2002). However, to examine how the type of the consumer demand forecast affects the quality of the planned orders, we describe a retail-specific model in detail here.

We consider a setting, in which each store can place product-specific orders only on predetermined order dates, which can vary across different stores and products. Figure 1 provides a visual summary on how the DC and the stores interact. The order quantity of an individual product in an individual store is determined by a periodic-review ordering policy, and thus planned orders for DC can be obtained by simulating the future replenishment orders of each SKU separately and then aggregating the simulation results over all examined SKUs.

3.1 *Computing planned orders with point-estimate demand forecasts*

Assume that there are $i \in \{1, \ldots, \bar{i}\}$ products and $j \in \{1, \ldots, \bar{j}\}$ stores (see Table 2 for summary of notation). Each SKU thus corresponds to a tuple $(i, j) \in \{1, \ldots, \bar{i}\} \times \{1, \ldots, \bar{j}\}$. We use S to denote a particular subset of all SKUs, i.e. $S \subseteq \{1, \ldots, \bar{i}\} \times \{1, \ldots \bar{j}\}$.

Table 2. Summary of the notation used in this paper.

Symbol	Description
$i \in \{1, \ldots, \bar{i}\}$	Product index
$j \in \{1, \ldots, \bar{j}\}$	Store index
$t \in T = \{1, \ldots, \bar{t}\}$	Time index (in days)
$S \subseteq \{1, \ldots, \bar{i}\} \times \{1, \ldots \bar{j}\}$	A subset of all SKUs
Inventory dynamics	
$X_{ij}(t)$	The demand of product i at store j on day t
$x_{ij}(t)$	Observed demand
$B_{ij}(t)$	Inventory level
$D_{ij}(t)$	Deliveries
$b^0_{ij}(t)$	Initial inventory level
$Q_{ij}(t)$	Ordered quantity
Parameters	
O	Set of order dates
c_{ij}	Case pack size of orders
k	Number of cases in a delivery
$l_{ij}(t)$	Lead time
$\tau_{ij}(t)$	Set of days in review and lead time period
m_{ij}	Minimum inventory requirement
$\hat{X}_{ij}(t)$	Point-estimate demand forecast
Sums of orders	
$Q_S(t)$	Total number of orders for group of SKUs S on day t
$Q^1_S(t)$	Forecasted total number of orders (based on point-estimate demand forecast)
$Q^2_S(t)$	Forecasted total number of orders (based on distribution demand forecast)
$Q^A_S(t)$	Total number of orders (actual replenishment ordering outcome)

The demand of product i at store j on day $t \in T = \{1, \ldots, \bar{t}\}$ is denoted by $X_{ij}(t)$. This demand can be satisfied with inventory level $B_{ij}(t)$ in the morning of day t as well as with deliveries $D_{ij}(t)$, which are assumed to arrive before the daily demand realises. Hence, the dynamics of the inventory level $B_{ij}(t)$ over time are given by

$$B_{ij}(t+1) = \max\left\{0, B_{ij}(t) + D_{ij}(t) - X_{ij}(t)\right\}, t \in T \tag{1}$$

where $B_{ij}(1) = b^0_{ij}$ is some fixed initial inventory.

In this paper, we assume that the stores use a periodic-review dynamic reorder point policy, where each replenishment order is a function of delivery schedule, minimum inventory requirement, case pack size and demand forecasts. Specifically, assume that store j can place orders for product i on a set of pre-specified days $O_{ij} \subset T$. The order quantity placed on day $t \in T$ is denoted by $Q_{ij}(t)$ and it is equal to zero if day t is not one of the order dates (i.e. $t \notin O_{ij}$). The case pack sizes $c_{ij} \in \mathbb{N}$ can vary among products and stores, and hence the order quantity satisfies $Q_{ij}(t) \in \{k \cdot c_{ij} | k \in \mathbb{N}\}$. Furthermore, the deterministic lead time for product i at store j is $l_{ij}(t) \in \mathbb{N}$ when ordered on day $t \in O_{ij}$. It is assumed that the delivery date of the current order date is always after the delivery date of the previous order date, i.e. that there are no order crossovers. This is a realistic assumption because violation of it would make it beneficial to refrain from ordering on the first-order date: the second order date would guarantee delivery at the same time or earlier, and this ordering decision could be based on more up-to-date information about the inventory level. With these definitions the deliveries of product i to store j in the beginning of day $t \in T$ are given by

$$D_{ij}(t) = \sum_{\substack{t' \in O_{ij} \\ t' + l_{ij}(t') = t}} Q_{ij}(t'). \tag{2}$$

Decision on the order quantity $Q_{ij}(t)$ depends on (i) the inventory level on the morning of the day t, and (ii) the deliveries and forecasted demand on the *review and lead time period* (*R+L-period*), which extends from the order day t to one day before the delivery of the *next* possible order date. Formally, the R+L-period of order date $t \in O_{ji}$ is

$$\tau_{ij}(t) = \{t, \ldots, t' + l_{ij}(t') - 1\}, \text{ where } t' = \min\{t'' \in O_{ij} | t'' > t\}. \tag{3}$$

Given a minimum inventory requirement $m_{ij} \geq 0$, the policy is to reorder the minimum amount necessary to keep the inventory above level m_{ij} in the end of the R+L-period assuming that the forecasted demand realises, i.e.

$$Q_{ij}(t) = \min_{k \in \mathbb{N}} \left\{ k \cdot c_{ij} \middle| B_{ij}(t) + \sum_{\hat{t} \in \tau_{ij}(t)} D_{ij}(\hat{t}) - \sum_{\hat{t} \in \tau_{ij}(t)} \hat{X}_{ij}(\hat{t}) + k \cdot c_{ij} \geq m_{ij} \right\}, \tag{4}$$

where $\tau_{ij}(t)$ is given by (3) and $\hat{X}_{ij}(t)$ is a point-estimate forecast of demand. Intuitively, this means that an order is placed only if the forecasted demand is large enough to decrease inventory level below the minimum inventory requirement during the R+L-period. We consider each m_{ij} to be an exogenously set parameter, whose size depends on marketing considerations such as the service level target.

The DC can estimate planned orders, i.e. forecasts of future replenishment orders, as follows: for each SKU the inventory level of the first-order date is forecasted by applying Equation (1) for all dates in T up to the first-order date. When applying Equation (1), the unobserved demands $X_{ij}(t)$ are estimated with the point-estimate forecasts $\hat{X}_{ij}(t)$, which are mean estimates from historical sales data. Based on the forecasted inventory level of the first-order date, the order quantity is obtained from Equation (4). Then inventory levels are iteratively computed up to the second-order date, and the order calculation is repeated. This procedure is continued until all order dates in $O_{ij} \subset T$ have been processed. We refer to this forecasting approach as the *point-estimate model*.

3.2 *Extension to distribution demand forecasts*

In the distribution forecast-based extension we assume that each $X_{ij}(t)$ is a random variable that follows some non-degenerate distribution on non-negative integers. With this assumption, Equations (1)–(5) jointly define a multivariate stochastic process in which all uncertainties ultimately stem from the uncertain consumer demand. We refer to this approach as the *probabilistic model*.

To compute planned orders with the probabilistic model, we use Monte Carlo simulation to obtain samples of daily order sums. In one simulation iteration, demand outcomes are drawn for each SKU for each day from their corresponding demand distributions. Then, this series of demand outcomes is used to compute inventory levels and calculate orders as in the point-estimate model. This procedure is repeated for all SKUs for several hundred times to obtain an estimate for distribution of orders for each order date.

It is important to highlight that this formulation allows the demand to be seasonal. Specifically, the point-estimate model formulation allows the point-estimate demand to vary over time, while the probabilistic model allows the demand distributions to vary over time. This is necessary in retail, because demand usually varies across, e.g. weekdays (Ehrenthal, Honhon, and Woensel 2014).

4. Comparison of point-estimate and probabilistic models based on real sales data

In this section, we use real data to compare the quality of the order forecasts obtained with the two planned order models. The two models differ in the type of consumer demand information they use: the first one uses point-estimate forecasts whereas the latter uses distribution forecasts.

We use a data-set of retail sales of SKUs from a European grocery retailer to estimate both models and to compare them. The models are compared based on both the technical accuracy of the order forecasts, and the monetary value these order forecasts have in supporting decision-making at the DC.

4.1 *Sales data and consumer demand forecasts*

The data-set we received from the retailer consists of 9155 SKUs with daily sales data from 1 January 2014 to 30 September 2014 for each SKU. The products are non-perishable grocery items, e.g. canned soups, cereal and potato chips. We used the first 181 days (6 months) of sales data to estimate the consumer demand distributions (training set) and the remaining 92 days (3 months) to evaluate the replenishment order forecasts (test set). The mean demand of all SKUs is 0.2 units per day throughout the time span so their demand can be characterised as low and intermittent. Furthermore, based on summary plots of the data, the SKU's exhibit seasonality over weekdays, which, as noted by Ehrenthal, Honhon, and Woensel (2014), is fairly common in retail.

To produce planned order forecasts for the test period's days, we estimate forecast models for consumer demand using the data of training period for each SKU separately. For this estimation, the sales data of each SKU is interpreted as realisations $x_{ij}(t)$ of the random demand $X_{ij}(t)$. To accommodate seasonality over weekdays in the forecasts, we used generalised linear regression models with weekday binaries as independent variables. The models were implemented using R's standard libaries and the MASS library (Venables and Ripley 2002). It should be noted that we did not detect significant amounts of autocorrelation in the data. Specifically, less than 3% of the SKUs exhibited autocorrelation (of any order) larger than 0.20 in the test data. If significant autocorrelation was found, INAR(n) models could be used to model the consumer demand (see, e.g. Alwan and Weiß 2017).

Three different distribution assumptions were tested in the regression models: Poisson, Geometric and Negative Binomial. The goodness of fit of these models was determined with Pearson Chi-Square tests (see, e.g. Agresti 2007). According to the tests, the risk of model misspecification was less than 0.05 for 44.4% of the SKUs in the case of Poisson regression, 52.5% in the case of Negative Binomial regression and 56.5% in the case of Geometric regression. Hence, the Pearson Chi-Square tests seem to indicate that both the Negative Binomial and the Geometric models would provide a good fit with the data. However, we also ran all simulations (see Sections 4.4–4.6) with all three regression models and did not observe significant differences in results between the different distribution assumptions. For this reason, we report results only from the geometric regression models.

To produce point-estimate forecasts for the SKUs, we used a weekday-specific version of the Croston model with Syntetos–Boylan Approximation (Syntetos and Boylan 2005). In this weekday-specific version, each weekday was forecasted separately based on demand history from the same weekday. The models were estimated using R's tsintermittent-library (Kourentzes and Petropoulos 2014). Use of the Croston model for point-estimate forecasting of products with low and intermittent demand was motivated by the model's widespread use in the industry (Teunter, Syntetos, and Babai 2011) as well as its good performance in academic forecast model evaluations (Syntetos, Babai, and Gardner 2015). Furthermore, we used the tsintermittent library (Kourentzes and Petropoulos 2014) to perform the demand classification test developed by Syntetos, Boylan, and Croston (2005). According to the results, the Syntetos–Boylan Approximation should be preferred over the normal Croston model for all SKUs in the data-set. We also tested the standard version of the Syntetos–Boylan Approximation, but it turned out that better results were obtained with the weekday-specific version. Hence, we report results only from the weekday-specific Croston model with the Syntetos–Boylan Approximation.

The consumer demand forecasts and replenishment order forecasts were produced with a weekly rolling horizon. Specifically, in the beginning of the test period the consumer demand of the entire test period was forecasted for each SKU separately (based on models fitted with training data). Then, in the beginning of the second week, the training data and the first week's data were used to fit the models again. This procedure was repeated up until the last week of test data, and thus the training data grew by one week in each iteration. Order simulations were also ran each week for the entire remaining test period. The purpose of this was to simulate a realistic planning procedure, where the consumer demand forecasts are updated regularly and where the updated consumer demand forecasts affect the forecasts of the stores' replenishment orders.

4.2 *Creating a test set-up for planned order models*

Because we received only sales data from the case company, we set some parameters of the replenishment ordering process so that our simulation closely resembles the real replenishment ordering of the retail chain. The parameters of our simulation are described in the following paragraphs.

Replenishment orders are placed using the inventory policy described in Section 3 (see Equation (4)). The parameters for the policy were chosen as follows: for each SKU, the minimum inventory requirement m_{ij} was set to equal six times the SKU's mean demand from the training period (rounded to whole numbers), which, for our products with low and intermittent demand, resulted in minimum inventory requirement m_{ij} of either one or two units across the SKUs. The case pack size c_{ij} was set to equal 4 for all SKUs. The initial inventory level at the beginning of the simulation was set to equal two times the minimum requirement for each SKU, i.e. $b^0_{ij} = 2m_{ij}$. These choices were motivated by the need to create a realistic test setup: first, in retail, the case pack sizes are often larger than one (see e.g. Curşeu et al. 2009), and thus the stores are often constrained to order and stock more than they would need. Second, based on discussions with the case company we argue that it is typical that, for store appearance purposes, m_{ij} is set to be multiple times larger than the average demand.

The case company uses different delivery schedules for different SKUs in order to balance the daily workload at the DC. To model this, we split the 9155 SKUs into different delivery schedule groups. Table 3 shows the delivery schedules for each group. We split SKUs into different groups based on their running database ID so that many delivery frequency options (7, 3, 2 and 1 per week) were sufficiently represented in the experiment. While it is common that retail SKUs have an option to be replenished daily or even multiple times per day (see e.g. Ehrenthal, Honhon, and Woensel 2014), Kuhn and Sternbeck (2013) note that more sparse delivery schedules are also very common and that the schedules are often varied across product groups.

Table 3. Delivery schedules used in the simulation. The crosses indicate if an SKU in a group has an option for a delivery on a particular weekday. For each SKU on each day $l_{ij}(t) = 1$, i.e. lead time is always one day.

Group	Number of SKUs	Delivery dates within a week						
		Mon	Tue	Wed	Thu	Fri	Sat	Sun
S^1	2305	x	x	x	x	x	x	x
S^2	2200	x		x		x		
S^3	1100		x			x		
S^4	1100				x			x
S^5	350	x						
S^6	350		x					
S^7	350			x				
S^8	350				x			
S^9	350					x		
S^{10}	350						x	
S^{11}	350							x

4.3 *Computation details*

Order forecasts of point-estimate and probabilistic models, are calculated for the test set with the procedures described in Section 3. In the DC, the number of distinct replenishment orders on a day t is often of interest, because the number of replenishment orders determines, e.g. the need for picking capacity on a particular day. Moreover, the order forecasts are often used on aggregate level for a group of SKUs. Hence, we evaluate the forecast accuracies of the two planned order models for groups of SKUs. The number of distinct orders on day t for SKUs belonging to set $S \subseteq \{1, \dots, \bar{i}\} \times \{1, \dots \bar{j}\}$ is computed as follows:

$$Q_S(t) = \sum_{\substack{(i,j)\in S \\ Q_{ij}(t)>0}} 1. \tag{5}$$

We denote the forecasts of distinct orders obtained from the point-estimate model and the probabilistic model by $Q_S^1(t)$ and $Q_S^2(t)$, respectively. It is worth noting that $Q_S^1(t)$ is a point-estimate whereas $Q_S^2(t)$ is a distribution. Therefore, the forecast accuracy metrics deployed later make use of the mean of $Q_S^2(t)$.

The total number of actual orders, denoted by $Q_S^A(t)$, is computed through the same procedure as the point-estimate forecast $Q_S^1(t)$, with the exception that in Equation (1) the actual observed demand from the sales data, $x_{ij}(t)$, is used instead of point-estimate forecast $\hat{X}_{ij}(t)$.

Generating $Q_S^A(t)$ in this fashion effectively means that we are examining an automated replenishment ordering scenario, where daily order volumes are forecasted with two competing forecasting methods. Therefore, we do not examine the effect the store personnel might have on the ordering process. For analysis on store personnel intervention on automated replenishment ordering (see van Donselaar et al. 2010).

It should also be noted, that the simulations use consumer demand forecasts in two ways. First in modelling the inventory balance dynamics (see Equation (1)) and second in the replenishment ordering policies (see Equation (4)). In our simulations, we used the Croston models' point-estimate forecasts $\hat{X}_{ij}(t)$ as replenishment order policies' forecasts in all simulations. Therefore, the procedures for computing $Q_S^A(t)$, $Q_S^1(t)$ and $Q_S^2(t)$ differ only with respect to how the inventory balance dynamics are modelled: $Q_S^A(t)$ uses the actual demand of test data, $Q_S^1(t)$ uses the point-estimate forecasts and $Q_S^2(t)$ uses the distributions forecasted with the regression models.

Computations were carried out using the R-language on a standard laptop (Dual-core, 2.6 GhZ, 8GB). In the probabilistic model, the distributions $Q_S^2(t)$ were estimated based on 300 Monte Carlo iterations for each SKU. Computation time of one iteration depends on the delivery schedule through the number of order days in the simulated time series. For instance in group S^1, where ordering is possible every day, one replication for one SKU took approximately 2.0 ms.

4.4 *Added forecast accuracy from probabilistic order forecast*

Accuracy of forecast models is evaluated with metrics that compare the forecasted and the actual order sums across possible order dates $t \in O_{ij}$. The metrics are computed for the sums within the groups presented in Table 3, as well as for a group that

Table 4. Replenishment order means, MSEs and mean average percentage errors by delivery schedule group for point-estimate and probabilistic models. All forecast accuracy measurements are averages across order dates possible within the corresponding group $S^{(\cdot)}$. The last group includes all SKUs, i.e. $S^{all} = \{1, \ldots, \bar{i}\} \times \{1, \ldots \bar{j}\}$.

SKU group	Number of SKUs in group	Mean of Q_S^A	MSE(Q_S^1)	MSE(Q_S^2)	MAPE(Q_S^1) (%)	MAPE(Q_S^2) (%)
S^1	2305	47.10	572.00	75.00	29.9	15.9
S^2	2200	121.71	1840.76	133.13	27.6	7.9
S^3	1100	96.36	1510.84	386.03	35.3	16.9
S^4	1100	96.92	1543.52	225.63	34.0	15.0
S^5	350	59.67	79.83	68.84	12.7	12.1
S^6	350	47.83	119.67	36.09	21.1	11.2
S^7	350	51.58	144.58	75.34	19.9	13.8
S^8	350	37.50	228.25	93.68	32.2	30.1
S^9	350	36.58	259.92	84.59	40.3	23.2
S^{10}	350	52.25	169.25	34.25	18.2	9.4
S^{11}	350	69.08	444.25	84.72	22.7	11.4
S^{all}	9055	199.44	4241.68	522.24	20.5	10.3

contains all SKUs. Measurement in these groups is reasonable, because a retail DC rarely needs to forecast replenishment orders of individual SKUs, whereas it often needs to forecast replenishment orders in groups of different sizes.

Two different forecast accuracy metrics are used to compare the two order simulation models: mean squared error (MSE) and mean average percentage error (MAPE). See the Appendix 1 for details and a numerical example of computing these metrics.

It should be noted that even though the SKUs we examine have low and intermittent demand, the aggregate replenishment orders across a group of hundreds of SKUs is likely to be high. Indeed, in each group that we examine, the sum of daily replenishment orders is larger than zero on each order date, and the daily replenishment order sum averages in all examined groups are between 36 (group S^9) and 199 (total across all groups, S^{all}). Therefore, MAPE's issues with zero-observations and observations close to zero are not a concern here, and we can use MAPE to compare the models. If the groups' daily order sums would have exhibited low and intermittent observations, mean-based metrics introduced by Prestwich et al. (2014) could have been used.

Kolassa (2016) discusses the use of more advanced metrics for measuring the forecast accuracy of distributions. Unfortunately, applying these more advanced metrics in this setting is difficult, since they are not well defined for point-estimate forecasts. Hence, the metrics would not allow us to compare point-estimate forecasts to distribution forecasts, which is the main focus of this paper.

Results of the accuracy evaluation are presented in Table 4. As we can see, the probabilistic model outperforms the point-estimate model in all groups and also jointly across the entire data-set. The differences in forecast accuracies across the groups can be attributed to errors in consumer demand forecasts and to the fact that – within a group – errors to opposite directions per SKU can cancel each other out on the aggregate level.

4.5 *Added value of improved accuracy in workforce planning*

The probabilistic model offers a significant improvement in forecasting accuracy, but this improvement has value only if it helps to make better operational decisions. To estimate the value for operational decisions, we present an example where the two models are applied to support workforce planning at the DC's daily order picking. We use workforce planning as an example application of operational decision-making, since order volumes are the dominant factor in workforce planning and since the significant cost impact of warehouse picking is well established (Petersen 1999; De Koster, Le-Duc, and Roodbergen 2007). Furthermore, our case company saw workforce planning as an application area where improved forecasting accuracy could relatively directly result in cost savings by allowing more efficient use of human resources.

In our example, the working schedules for the regular staff responsible for handling orders at the DC have to be prepared in advance for a planning period of three months (the test period). If the orders coming in to the DC exceed, the handling capacity of the preplanned regular staff, then excess orders have to be handled by temporary workers at a higher cost.

Formally, on day $t = 1$ the DC decides on daily handling capacity produced by the regular staff, denoted by $h(t)$, for days within the three-month planning period, i.e. $t \in T' = \{1, \ldots, 91\}$. The unit cost of this capacity is $c_1 = 1$. If on the day t the

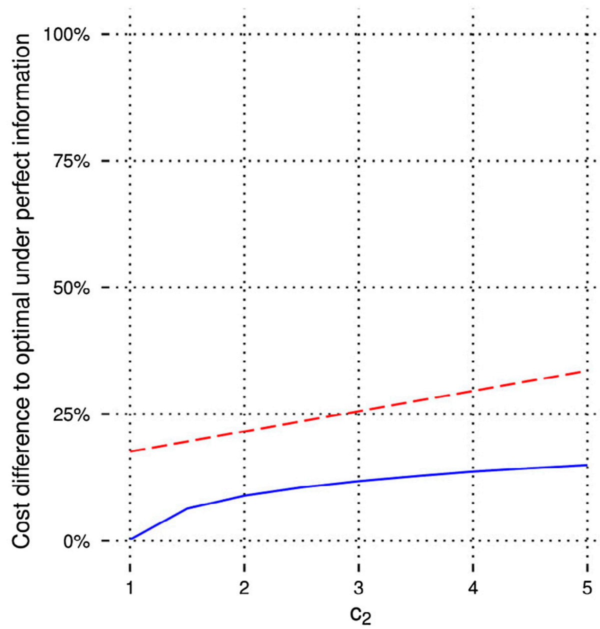

Figure 2. Costs differences as a function of c_2 relative to a case with perfect forecasts. The dashed line represents the point-estimate model and the solid line represents the probabilistic model.

number of distinct orders across all SKUs, $Q_{S^{all}}(t)$ (see, Equation (5)), exceeds the handling capacity (i.e. $Q_{S^{all}}(t) > h(t)$) then the excess orders $Q_{S^{all}}(t) - h(t)$ have to be handled by temporary staff, whose handling capacity has a higher unit cost $c_2 > c_1$. Hence, the DC seeks to minimise the total expected workforce cost over the planning period, i.e.

$$\min_{h(t),\ t\in T'} \mathbb{E}\left[\sum_{t\in T'} \left(c_1 h(t) + c_2 \max\{0, Q_{S^{all}}(t) - h(t)\}\right)\right]. \tag{6}$$

which is equivalent to

$$\sum_{t\in T'} \min_{h(t)} \mathbb{E}\left[\left(\max\{c_1 h(t), c_2 Q_{S^{all}}(t) - (c_1 - c_2)h(t)\}\right)\right]. \tag{7}$$

This corresponds to a sum of well-known newsvendor problems (see e.g. Porteus 1990), for which the use of standard differentiation techniques yields the optimal solution

$$h^*(t) = F^{-1}_{Q_{S^{all}}(t)}\left(\frac{c_2 - c_1}{c_2}\right), t \in T', \tag{8}$$

where $F^{-1}_{Q_{S^{all}}(t)}$ is the inverse cumulative distribution function of $Q_{S^{all}}(t)$. In case of the probabilistic model, $F^{-1}_{Q^2_{S^{all}}(t)}$ can be estimated from the samples of $F_{Q^2_{S^{all}}(t)}$ generated through Monte Carlo simulation. In case of the point-estimate model, $F^{-1}_{Q^1_{S^{all}}(t)}$ is a step function and hence the optimal solution becomes $h^*(t) = Q^1_{S^{all}}(t)$.

To see how the cost difference between the planned and the temporary capacity affects the total costs for both models, we conduct a sensitivity analysis over a range of possible values for c_2. Figure 2 shows both models' costs compared to a case, where the case company would have perfect forecasts for each t. As we can see, the total workforce costs are significantly lower when probabilistic order forecasts are used in planning, and the advantage of the probabilistic model over the point-estimate model persists irrespective of c_2.

4.6 *Added value of improved accuracy in inventory management*

As another example of estimating the value of improved forecasts for operational decisions, we apply the two planned order models to support inventory management at the DC. In this example, the retail stores are replenished from the DC with the

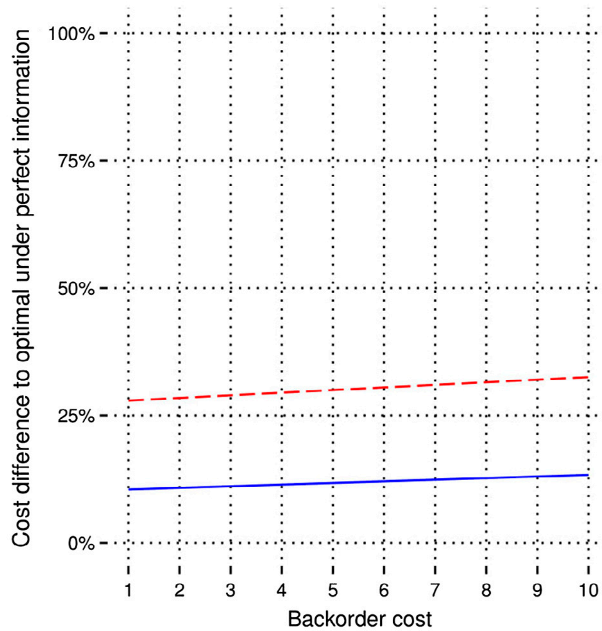

Figure 3. Cost differences as a function of the backorder cost relative to a case with perfect forecasts. The dashed line represents the point-estimate model and the solid line represents the probabilistic model. The inventory holding cost is 5.

delivery schedules and parameters described in Section 4.2. The DC, in turn, places replenishment orders to its suppliers in to satisfy the demand from the stores.

Inventory is managed at the DC on a product basis, and thus the examined 9155 SKUs are grouped by their product code (obtained from the case company). This results in 511 groups with 2 to 81 SKUs in each group. The DC can place replenishment orders for each product once a week with a lead time of 7 days, and the DC uses the same inventory policy the stores use to place replenishment orders to the DC (see Equation (4)).

DC's inventory policy is parameterised as follows: case pack size is one for all products, i.e. it is assumed that the DC orders case packs from the supplier and supplies the same case packs to the stores. The DC aims to maintain an inventory that is ten times larger than the forecasted mean daily demand, and for each case pack in inventory the DC faces an inventory holding cost per day. If on any day the demand exceeds the DCs inventory, the demand is backordered and the DC faces a backorder cost per case pack.

DC's inventory management was simulated over the test period, and realised costs were computed using both planned order models. Sensitivity analysis over the backorder cost was conducted, and the results are presented in Figure 3. As we can see, the better forecast accuracy of the probabilistic model also translates to lower inventory management costs irrespective of the ratio between the holding costs and the backorder costs.

5. Discussion

The results presented in Section 4 show that the probabilistic model clearly outperforms the point-estimate model. The reason for this can be found by taking a closer look at the dynamics of retail's replenishment ordering.

By Equation (4), the order policy can be interpreted as a function $g(\cdot)$ of the order date's inventory level, which gives the replenishment order $Q_{ij}(t) = g(B_{ij}(t))$. Thus, using the expected replenishment order $\mathbb{E}[Q_{ij}(t)] = \mathbb{E}[g(B_{ij}(t))]$ as the order forecast calls for applying the order policy $g(\cdot)$ over the distribution of $B_{ij}(t)$. However, with point-estimate consumer demand forecasts, the distribution cannot be inferred, and instead the point-estimate model uses $\mathbb{E}[B_{ij}(t)]$ as the input for $g(\cdot)$. Effectively, the point-estimate model thus assumes that $\mathbb{E}[g(B_{ij}(t))] = g(\mathbb{E}[B_{ij}(t)])$, which would hold if $g(\cdot)$ was a linear function. However, as we can see from the Equation (4), the applied order policy is not linear in inventory level, but instead a step function. In fact, any inventory policy $g(\cdot)$ designed for integer-valued inventory with case pack size constraints is a step function. Hence, in general $\mathbb{E}[g(B_{ij}(t))] \neq g(\mathbb{E}[B_{ij}(t)])$ which means that the expected order quantity is not equal to the order quantity resulting from the expected inventory level. Therefore, using point-estimate forecasts as a

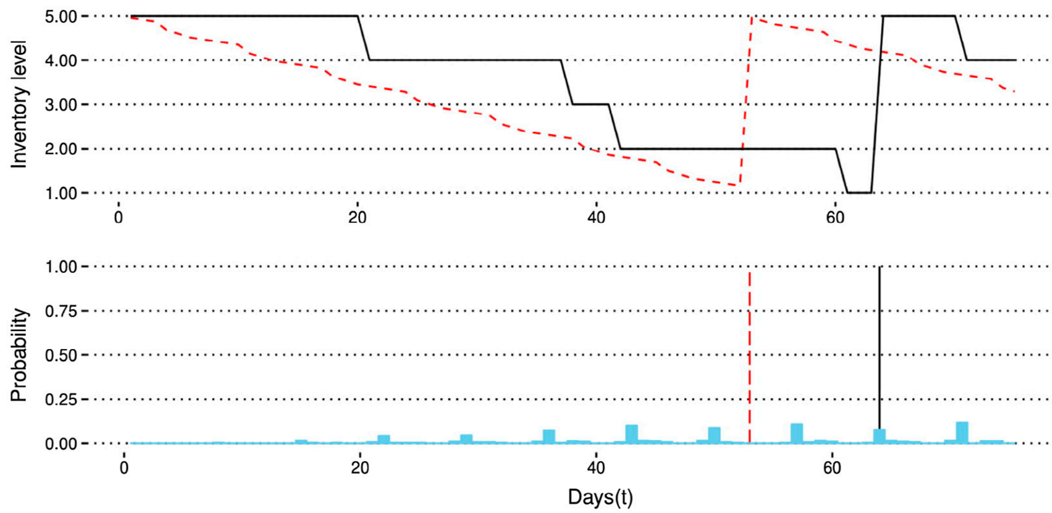

Figure 4. The top graph illustrates the point-estimate forecast for the SKU's inventory level (dashed line) as well as the actual realised inventory level (solid line). The bottom graph compares the timing of the corresponding replenishment orders to the order probabilities obtained from the probabilistic model (bars).

basis for computing planned orders can be expected to yield sufficiently accurate forecasts of replenishment orders only if the simulated inventory levels of order dates are very close to actual inventory levels of those dates.

With point-estimate forecasts the estimated change in daily inventory is always equal to the day's forecasted mean demand. Consequently, the point-estimate model assumes that the need to order arises after a large number of days with mean demand have passed, i.e. after the sum of point-estimate forecasts is large enough to cause the forecasted inventory level to fall under the minimum inventory requirement. This kind of model for inventory dynamics is quite different from the reality of products with low and intermittent demand: inventory level movements under low and intermittent demand are characterised by long periods of zero realisations accompanied by irregularly appearing demands that are relatively large when compared to the long-term mean (which – due to the large number of zero realisations – is often close to zero). Upper graph in Figure 4 illustrates the difference in inventory dynamics between the point-estimate model and the actual orders. As we can see, the inventory level movements forecasted by the point-estimate model (dashed line) are regular and driven by daily flows of forecasted means, whereas the actual inventory level movements (solid line) are stepwise.

The lower graph of Figure 4 compares the estimates of forecasted order probabilities for the same SKU as in the upper graph. As we can see, the probabilistic model predicts that orders are possible on a range of dates with relatively low probabilities for each day, whereas the point-estimate model predicts that the order is certain to occur on a specific date. Hence, the planned order forecasts for the same SKU are very different with the two models. From the results in Table 4, we observe that the mean estimates of replenishment orders forecasted with the probabilistic model are clearly more accurate than their point-estimate counterparts.

The point-estimate model has potential to work better when the demand is not low and intermittent. Specifically, if the demand of an SKU is symmetric with some relatively high mean and relatively low variance, then the estimates of future inventory levels made with point-estimate forecasts might be accurate even if the demand realisations of individual days would not exactly equal the point-estimate means. This can happen because, for symmetric distributions, upward deviations on one day are likely to be cancelled out by downward deviations on other days (or vice versa).

From the last line in Table 4, we observe that the forecast errors in individual groups cancel each other out somewhat when the aggregate across all SKUs is examined. Indeed, the point-estimate model benefits from the summation so much that the average MAPE value for the total sum is lower than majority of the group-specific MAPE values. Indeed, the point-estimate model's MAPE value of 21% can be acceptable in many planning problems. This implies that the point-estimate model can be a useful tool if it is applied to a decision problem, which involves a large number of SKUs with varying delivery schedules. However, as we can see from Table 4 the probabilistic model still outperformed the point-estimate model by a clear margin even in this case. Furthermore, from Sections 4.5 and 4.6 we observe that, in the application examples, the difference in forecast accuracy also translated to differences in costs in favour of the probabilistic model.

6. Conclusions

In retail, the distribution centres need to forecast the replenishment orders the stores generate on a daily basis. Based on earlier results, we argue in favour of preferring planned orders over more straightforward forecasting approaches, such as time series models, which are based only on POS data and/or order history.

The focus of this paper was to explore how accurately planned orders predict the replenishment orders of products with low and intermittent demand, and how the accuracy of the planned orders could be increased. To achieve this, we described two planned order models: one that uses point-estimates to model consumer demand (point-estimate model) and another that uses distributions (probabilistic model). The probabilistic model's forecast accuracy was compared to the accuracy of the point-estimate model in a replenishment ordering simulation conducted with real sales data of low and intermittent retail demand. In this comparison, we observed that the probabilistic model outperformed the point-estimate model in all test cases. The models were also compared in two applications, where forecasts from both models were used to support inventory management and workforce capacity planning at the DC. Here, we observed that using the probabilistic model yielded substantially lower costs in both application examples.

Based on our results, we conclude that planned orders are useful in DC's order forecasting if appropriate steps are taken in the modelling. Specifically, for products with low and intermittent demand a probability model of consumer demand should be preferred over point-estimate models. These observations contribute to discussion about forecasting with planned orders: Earlier Zhao, Xie, and Leung (2002) and Byrne and Heavey (2006) have concluded that cost savings can be achieved if planned orders are used as forecasts. We contributed by presenting a detailed model for computing planned orders and by pointing out that distribution forecasts of consumer demand should be preferred if planned orders are computed under low and intermittent consumer demand.

Our results are in line with earlier findings on the potential added value of sharing information in the supply chain (see, e.g. Ketzenberg and Ferguson 2008; Kiesmüller and Broekmeulen 2010; Zhang and Cheung 2011). To fulfil this potential in retail, the developers of automated store ordering software should consider developing support for modelling consumer demand with distributions. Once capabilities for computing distribution forecasts exist, planned order modules that make use of distribution forecasts should be developed. Encouragingly, our forecast accuracy measurements show that even a simple probability model can provide reasonably accurate order forecasts. These results lend support to literature that advocates rigorous modelling of demand uncertainty in operations management (e.g. Feng, Rao, and Raturi 2011; Käki, Salo, and Talluri 2013; Chapados 2014; Käki et al. 2015).

This research opens several avenues for further research. First, it highlights the need to develop distribution-based forecast models that address various, more complicated, demand scenarios in retail. Chapados (2014) has developed a general framework for developing negative binomial time series models for demand forecasting. It can be used as a starting point for developing more specific applications. Second, our results call for more research on measuring the benefits of different forecasting techniques: based on our results, we can argue that the DC's inventory management and workforce planning could benefit from using planned orders whereas Jonsson and Mattsson (2013) can, based on their results, argue that planned orders have a tendency to increase supplier's inventory levels. These two views can result in conflicting recommendations and thus more research is required in order to establish what are the net benefits of using different forecasting techniques and what factors drive the differences in recommendations. Third, distribution forecasts of orders present an opportunity to study risk management in DC's workforce planning. For example, if large underestimations of order volumes have high costs, forecast distributions can be used to assess the likelihood of these situations.

Acknowledgements
We would like to thank the four reviewers and the associate editor for their comments and suggestions that helped us to improve this paper.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This paper was supported financially by the HSE Foundation and the Aalto University School of Business.

ORCID
Ville Sillanpää http://orcid.org/0000-0003-1909-7941

References

Agrawal, N., and S. A. Smith. 1996. "Estimating Negative Binomial Demand for Retail Inventory Management with Unobservable Lost Sales." *Naval Research Logistics* 43 (6): 839–861.

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. New Jersey: John Wiley.

Alftan, A., R. Kaipia, L. Loikkanen, and K. Spens. 2015. "Centralised Grocery Supply Chain Planning: Improved Exception Management." *International Journal of Physical Distribution & Logistics Management* 45 (3): 237–259.

Al-Osh, M., and A. A. Alzaid. 1987. "First-order Integer-valued Autoregressive (INAR (1)) Process." *Journal of Time Series Analysis* 8 (3): 261–275.

Alwan, L. C., and C. H. Weiß. 2017. "Inar Implementation of Newsvendor Model for Serially Dependent Demand Counts." *International Journal of Production Research* 55 (4): 1085–1099.

Aviv, Y. 2002. "Gaining Benefits from Joint Forecasting and Replenishment Processes: The Case of Auto-correlated Demand." *Manufacturing & Service Operations Management* 4 (1): 55–74.

Aviv, Y. 2007. "On the Benefits of Collaborative Forecasting Partnerships between Retailers and Manufacturers." *Management Science* 53 (5): 777–794.

Boylan, J. E., and A. A. Syntetos. 2010. "Spare Parts Management: A Review of Forecasting Research and Extensions." *IMA journal of management mathematics* 21 (3): 227–237.

Byrne, P., and C. Heavey. 2006. "The Impact of Information Sharing and Forecasting in Capacitated Industrial Supply Chains: A Case Study." *International Journal of Production Economics* 103 (1): 420–437.

Campo, K., E. Gijsbrechts, and P. Nisol. 2000. "Towards Understanding Consumer Response to Stock-outs." *Journal of Retailing* 76 (2): 219–242.

Cannella, S., and E. Ciancimino. 2010. "On the Bullwhip Avoidance Phase: Supply Chain Collaboration and Order Smoothing." *International Journal of Production Research* 48 (22): 6739–6776.

Chapados, N. 2014. "Effective Bayesian Modeling of Groups of Related Count Time Series." *Proceedings of The 31st International Conference on Machine Learning*, Montreal, 1395–1403.

Croston, J. D. 1972. "Forecasting and Stock Control for Intermittent Demands." *Operational Research Quarterly* 23: 289–303.

Croston, J. 1974. "Stock Levels for Slow-moving Items." *Journal of the Operational Research Society* 25 (1): 123–130.

Curşeu, A., T. Van Woensel, J. Fransoo, K. Van Donselaar, and R. Broekmeulen. 2009. "Modelling Handling Operations in Grocery Retail Stores: An Empirical Analysis." *Journal of the Operational Research Society* 60 (2): 200–214.

De Koster, R., T. Le-Duc, and K. J. Roodbergen. 2007. "Design and Control of Warehouse Order Picking: A Literature Review." *European Journal of Operational Research* 182 (2): 481–501.

Dolgui, A., and M. Pashkevich. 2008. "On the Performance of Binomial and Beta-binomial Models of Demand Forecasting for Multiple Slow-moving Inventory Items." *Computers & Operations Research* 35 (3): 893–905.

Dolgui, A., A. Pashkevich, and M. Pashkevich. 2004. "Modeling Demand for Inventory Management of Slow-moving Items in Case of Reporting Errors." *IFAC Proceedings Volumes* 37 (4): 27–32.

Dolgui, A., A. Pashkevich, and M. Pashkevich. 2005. "Bayesian Approach to Modelling of Quasi-periodic Intermittent Demand." *IFAC Proceedings Volumes* 38 (1): 343–348.

Ehrenthal, J., D. Honhon, and T. V. Woensel. 2014. "Demand Seasonality in Retail Inventory Management." *European Journal of Operational Research* 238 (2): 527–539.

Emmelhainz, M. A., J. R. Stock, and L. W. Emmelhainz. 1991. "Consumer Responses to Retail Stock-outs." *Journal of Retailing* 67 (2): 138–148.

Esparcia-Alcazar, A., E. Alfaro-Cid, P. Garcia-Sanchez, A. Martinez-Garcia, J. Merelo, and K. Sharman 2010. "An Evolutionary Approach to Integrated Inventory and Routing Management in a Real World Case." In *2010 IEEE Congress on Evolutionary Computation (CEC)*, Barcelona, 1–7.

Feng, K., U. S. Rao, and A. Raturi. 2011. "Setting Planned Orders in Master Production Scheduling under Demand Uncertainty." *International Journal of Production Research* 49 (13): 4007–4025.

Fernie, J., F. Pfab, and C. Marchant. 2000. "Retail Grocery Logistics in the UK." *The International Journal of Logistics Management* 11: 83–90.

Fernie, J., L. Sparks, and A. C. McKinnon. 2010. "Retail Logistics in the UK: Past, Present and Future." *International Journal of Retail & Distribution Management* 38: 894–914.

Gaur, V., and M. L. Fisher. 2004. "A Periodic Inventory Routing Problem at a Supermarket Chain." *Operations Research* 52 (6): 813–822.

Giloni, A., C. Hurvich, and S. Seshadri. 2014. "Forecasting and Information Sharing in Supply Chains under Arma Demand." *IIE Transactions* 46 (1): 35–54.

Holzapfel, A., A. Hübner, H. Kuhn, and M. G. Sternbeck. 2015. "Delivery Pattern and Transportation Planning in Grocery Retailing." *European Journal of Operational Research* 252 (1): 54–68.

Jin, Y., B. D. Williams, T. Tokar, M. A. Waller, et al. 2015. "Forecasting with Temporally Aggregated Demand Signals in a Retail Supply Chain." *Journal of Business Logistics* 36 (2): 199–211.

Jonsson, P., and S.-A. Mattsson. 2013. "The Value of Sharing Planning Information in Supply Chains." *International Journal of Physical Distribution & Logistics Management* 43 (4): 282–299.

Käki, A., J. Liesiö, A. Salo, and S. Talluri. 2015. “Newsvendor Decisions under Supply Uncertainty.” *International Journal of Production Research* 53 (5): 1544–1560.

Käki, A., A. Salo, and S. Talluri. 2013. “Impact of the Shape of Demand Distribution in Decision Models for Operations Management.” *Computers in Industry* 64 (7): 765–775.

Ketzenberg, M., and M. E. Ferguson. 2008. “Managing Slow-moving Perishables in the Grocery Industry.” *Production and Operations Management* 17 (5): 513–521.

Kiesmüller, G., and R. Broekmeulen. 2010. “The Benefit of VMI Strategies in a Stochastic Multi-product Serial Two Echelon System.” *Computers & Operations Research* 37 (2): 406–416.

Koh, S., S. Saad, and M. Jones. 2002. “Uncertainty under MRP-planned Manufacture: Review and Categorization.” *International journal of production research* 40 (10): 2399–2421.

Kolassa, S. 2016. “Evaluating Predictive Count Data Distributions in Retail Sales Forecasting.” *International Journal of Forecasting* 32 (3): 788–803.

Kourentzes, N., and F. Petropoulos. 2014. “tsintermittent: Intermittent Time Series Forecasting.” R package. https://CRAN.R-project.org/package=tsintermittent.

Kuhn, H., and M. Sternbeck. 2013. “Integrative Retail Logistics: An Exploratory Study.” *Operations Management Research* 6 (1–2): 2–18.

Levén, E., and A. Segerstedt. 2004. “Inventory Control with a Modified Croston Procedure and Erlang Distribution.” *International journal of production economics* 90 (3): 361–367.

Marklund, J. 2002. “Centralized Inventory Control in a Two-level Distribution System with Poisson Demand.” *Naval Research Logistics (NRL)* 49 (8): 798–822.

McKenzie, E. 1985. “Some Simple Models for Discrete Variate Time Series.” *JAWRA Journal of the American Water Resources Association* 21 (4): 645–650.

Montanari, R., G. Ferretti, M. Rinaldi, and E. Bottani. 2015. “Investigating the Demand Propagation in EOQ Supply Networks Using a Probabilistic Model.” *International Journal of Production Research* 53 (5): 1307–1324.

Monthatipkul, C., and P. Yenradee. 2008. “Inventory/distribution Control System in a One-warehouse/Multi-retailer Supply Chain.” *International Journal of Production Economics* 114 (1): 119–133.

Nikolopoulos, K. I., M. Z. Babai, and K. Bozos. 2016. “Forecasting Supply Chain Sporadic Demand with Nearest Neighbor Approaches.” *International Journal of Production Economics* 177: 139–148.

Nikolopoulos, K., A. A. Syntetos, J. E. Boylan, F. Petropoulos, and V. Assimakopoulos 2011. “An Aggregate-disaggregate Intermittent Demand Approach (ADIDA) to Forecasting: An Empirical Proposition and Analysis.” *Journal of the Operational Research Society* 62: 544–554.

Orlicky, J. 1975. *Material Requirements Planning: The New Way of Life in Production and Inventory Management*. McGraw-Hill. https://isbndb.com/book/9780070477087.

Petersen, C. G. 1999. “The Impact of Routing and Storage Policies on Warehouse Efficiency.” *International Journal of Operations & Production Management* 19 (10): 1053–1064.

Porteus, E. L. 1990. “Stochastic Inventory Theory.” *Handbooks in operations research and management science* 2: 605–652.

Prestwich, S., R. Rossi, S. Armagan Tarim, and B. Hnich. 2014. “Mean-based Error Measures for Intermittent Demand Forecasting.” *International Journal of Production Research* 52 (22): 6782–6791.

Rostami-Tabar, B., M. Z. Babai, A. Syntetos, and Y. Ducq. 2013. “Demand Forecasting by Temporal Aggregation.” *Naval Research Logistics (NRL)* 60 (6): 479–498.

Sari, K. 2008. “On the Benefits of CPFR and VMI: A Comparative Simulation Study.” *International Journal of Production Economics* 113 (2): 575–586.

Schary, P. B., and M. Christopher. 1979. “Anatomy of a Stock-out.” *Journal of Retailing* 55 (2): 59–70.

Småros, J. 2007. “Forecasting Collaboration in the European Grocery Sector: Observations from a Case Study.” *Journal of Operations Management* 25 (3): 702–716.

Småros, J., A. Angerer, J. Fernie, L. Toktay, and G. Zotteri 2004. “Logistics Processes of European Grocery Retailers.” Helsinki University of Technology Working Paper.

Snyder, R. D., J. K. Ord, and A. Beaumont. 2012. “Forecasting the Intermittent Demand for Slow-moving Inventories: A Modelling Approach.” *International Journal of Forecasting* 28 (2): 485–496.

Sternbeck, M. G., and H. Kuhn. 2014. “An Integrative Approach to Determine Store Delivery Patterns in Grocery Retailing.” *Transportation Research Part E: Logistics and Transportation Review* 70: 205–224.

Syntetos, A. A., M. Z. Babai, and N. Altay. 2012. “On the Demand Distributions of Spare Parts.” *International Journal of Production Research* 50 (8): 2101–2117.

Syntetos, A. A., M. Z. Babai, and E. S. Gardner. 2015. “Forecasting Intermittent Inventory Demands: Simple Parametric Methods vs. Bootstrapping.” *Journal of Business Research* 68 (8): 1746–1752.

Syntetos, A. A., and J. E. Boylan. 2001. “On the Bias of Intermittent Demand Estimates.” *International Journal of Production Economics* 71 (1): 457–466.

Syntetos, A. A., and J. E. Boylan. 2005. “The Accuracy of Intermittent Demand Estimates.” *International Journal of forecasting* 21 (2): 303–314.

Syntetos, A. A., J. E. Boylan, and J. Croston. 2005. “On the Categorization of Demand Patterns.” *Journal of the Operational Research Society* 56 (5): 495–503.

Syntetos, A. A., J. E. Boylan, and S. M. Disney. 2009. "Forecasting for Inventory Planning: A 50-year Review." *Journal of the Operational Research Society* s149–s160.

Syntetos, A., D. Lengu, and M. Z. Babai. 2013. "A Note on the Demand Distributions of Spare Parts." *International Journal of Production Research* 51 (21): 6356–6358.

Teunter, R. H., A. A. Syntetos, and M. Z. Babai. 2011. "Intermittent Demand: Linking Forecasting to Inventory Obsolescence." *European Journal of Operational Research* 214 (3): 606–615.

Tiwari, V., and S. Gavirneni. 2007. "ASP, the Art and Science of Practice: Recoupling Inventory Control Research and Practice: Guidelines For Achieving Synergy." *Interfaces* 37 (2): 176–186.

Trapero, J. R., N. Kourentzes, and R. Fildes. 2015. "Identification of Sales Forecasting Models." *Journal of the Operational Research Society* 66 (2): 299–307.

van Donselaar, K. H., V. Gaur, T. van Woensel, R. A. C. M. Broekmeulen, and J. C. Fransoo. 2010. "Ordering Behavior in Retail Stores and Implications for Automated Replenishment." *Management Science* 56 (5): 766–784.

van Donselaar, K., T. van Woensel, R. Broekmeulen, and J. Fransoo. 2006. "Inventory Control of Perishables in Supermarkets." *International Journal of Production Economics* 104 (2): 462–472.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4th ed. New York: Springer.

Vollmann, T. E., W. L. Berry, and D. C. Whybark. 1997. *Manufacturing Planning and Control Systems*. New York: Irwin/McGraw-Hill.

Wagner, H. M., and T. M. Whitin. 1958. "Dynamic Version of the Economic Lot Size Model." *Management Science* 5 (1): 89–96.

Waller, M., M. E. Johnson, and T. Davis. 1999. "Vendor-managed Inventory in the Retail Supply Chain." *Journal of Business Logistics* 20: 183–204.

Willemain, T. R., C. N. Smart, and H. F. Schwarz. 2004. "A New Approach to Forecasting Intermittent Demand for Service Parts Inventories." *International Journal of forecasting* 20 (3): 375–387.

Williams, T. 1984. "Stock Control with Sporadic and Slow-moving Demand." *Journal of the Operational Research Society* 35: 939–948.

Williams, B. D., and M. A. Waller. 2011. "Estimating a Retailer's Base Stock Level: An Optimal Distribution Center Order Forecast Policy." *Journal of the Operational Research Society* 62 (4): 662–666.

Williams, B. D., M. A. Waller, S. Ahire, and G. D. Ferrier. 2014. "Predicting Retailer Orders with POS and Order Data: The Inventory Balance Effect." *European Journal of Operational Research* 232 (3): 593–600.

Xu, K., and M. T. Leung. 2009. "Stocking Policy in a Two-party Vendor Managed Channel with Space Restrictions." *International Journal of Production Economics* 117 (2): 271–285.

Zhang, S. H., and K. L. Cheung. 2011. "The Impact of Information Sharing and Advance Order Information on a Supply Chain with Balanced Ordering." *Production and Operations Management* 20 (2): 253–267.

Zhang, M., G. Nie, Z. He, and X. Hou. 2014. "The Poisson INAR (1) One-sided EWMA Chart with Estimated Parameters." *International Journal of Production Research* 52 (18): 5415–5431.

Zhang, Y., C. Zhang, and Y. Liu. 2016. "An AHP-based Scheme for Sales Forecasting in the Fashion Industry." In *Analytical Modeling Research in Fashion Business*, edited by T. -M. Choi, 251–267. New York: Springer.

Zhao, X., J. Xie, and J. Leung. 2002. "The Impact of Forecasting Model Selection on the Value of Information Sharing in a Supply Chain." *European Journal of Operational Research* 142 (2): 321–344.

Zinn, W., and P. C. Liu. 2001. "Consumer Response to Retail Stockouts." *Journal of Business Logistics* 22 (1): 49–71.

Zinn, W., and P. C. Liu. 2008. "A Comparison of Actual and Intended Consumer Behavior in Response to Retail Stockouts." *Journal of Business Logistics* 29 (2): 141–159.

Appendix 1.

A.1 *Computing the forecast accuracy metrics*

The average mean squared error of forecasts $Q_S(t)$ and actual outcomes $Q_S^A(t)$ for a group of SKUs S is given by

$$\mathrm{MSE}(Q_S) = \frac{1}{|O_S|} \sum_{t \in O_S} [Q_S^A(t) - \mathrm{mean}(Q_S(t))]^2, \tag{A1}$$

where $|O_S|$ denotes the number of order dates of all SKUs that belong to group S (i.e. $O_S = \cup_{(i,j)\in S} O_{ij}$). The mean absolute percentage error $\mathrm{MAPE}(Q_S)$ is given by

$$\mathrm{MAPE}(Q_S) = \frac{100\%}{|O_S|} \sum_{t \in O_S} \frac{|Q_S^A(t) - \mathrm{mean}(Q_S(t))|}{Q_S^A(t)}. \tag{A2}$$

As an example of computing both metrics, we consider the SKU group S^9 (see Table 3). In this group, for each SKU, there is an option to place a replenishment order every Friday. This translates to 12 order dates during the test period (i.e. $O_{S^9} = \{4, 11, 18, 25, 32, 39, 46, 53, 60, 67, 74, 81\}$). Using Equation (5) the quantities $Q_S^A(t)$, $Q_S^1(t)$ and $Q_S^2(t)$ are computed for each $t \in O_{S^9}$, and then Equations (A1) and (A2) are applied to compute MSE and MAPE. Table A1 presents the details of these computations. As we can see, both metrics's means over order dates equal the means reported in Table 4 for the SKU-group S^9. Moreover, even for this group with the lowest average number of replenishment orders per day, the daily order quantity varies between 6 and 45, i.e. there are no order dates with zero replenishment orders. Therefore MAPE can be reliably used to compare the two forecast models.

Table A1. Forecast accuracy metrics computed for delivery schedule group S^9.

t	$Q^A_{S^9}(t)$	$Q^1_{S^9}(t)$	$\text{mean}(Q^2_{S^9}(t))$	$\left[Q^A_{S^9}(t) - Q^1_{S^9}(t)\right]^2$	$\left[Q^A_{S^9}(t) - \text{mean}(Q^2_{S^9}(t))\right]^2$	$\frac{\lvert Q^A_{S^9}(t) - Q^1_{S^9}(t)\rvert}{Q^A_{S^9}(t)}$	$\frac{\lvert Q^A_{S^9}(t) - \text{mean}(Q^2_{S^9}(t))\rvert}{Q^A_{S^9}(t)}$
4	6	1	8.06	25.00	4.24	83.3 %	34.3 %
11	43	70	55.48	729.00	155.75	62.8 %	29.0 %
18	42	72	53.75	900.00	138.06	71.4 %	28.0 %
25	42	49	51.43	49.00	88.92	16.7 %	22.5 %
32	30	49	48.62	361.00	346.70	63.3 %	62.1 %
39	41	41	46.43	0.00	29.48	0.0 %	13.2 %
46	35	60	44.25	625.00	85.56	71.4 %	26.4 %
53	35	46	43.47	121.00	71.74	31.4 %	24.2 %
60	44	56	42.55	144.00	2.10	27.3 %	3.3 %
67	33	43	42.19	100.00	84.46	30.3 %	27.8 %
74	45	41	42.29	16.00	7.34	8.9 %	6.0 %
81	43	50	42.26	49.00	0.55	16.3 %	1.7 %
				$\text{MSE}(Q^1_{S^9})$	$\text{MSE}(Q^2_{S^9})$	$\text{MAPE}(Q^1_{S^9})$	$\text{MAPE}(Q^2_{S^9})$
				259.92	84.59	40.3 %	23.2 %