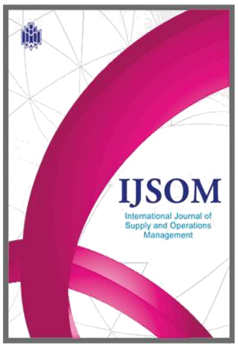
International Journal of Supply and Operations Management

IJSOM

February 2023, Volume 10, Issue 1, pp. 89-104
ISSN-Print: 2383-1359
ISSN-Online: 2383-2525
www.ijsom.com



Demand Models for Supermarket Demand Forecasting

Ulrich Kerzel [a, *]

[a] *Department of IT and technology, IU international University of applied sciences, Erfurt, Germany*

Abstract

Model-based approaches remain an important option for modelling customer demand. While this approach allows to analyse demand using a model based on theoretical arguments, each choice of model is associated with specific assumptions under which this model is valid. Customer demand in retail is typically modelled as a Poison-type process, in particular using a negative binomial distribution. Poisson-type processes are associated with an exponential inter-arrival time that describes the probability distribution between subsequent events. Using a public dataset from a large supermarket, the analysis of the data shows that while the general assumption of a Poisson-process is reasonable, the purchasing behaviour strongly depends on the type of product. Additionally, customers in this supermarket show a strong preference for a weekly shopping trip.

Keywords: Stochastic demand; Demand model; Poisson; Negative binomial.

1. Introduction

One of the core operational aspects of retailers is to ensure that sufficient inventory is maintained to satisfy customer demand. Conceptually, we can split this task into two aspects: First, we need to estimate the future customer demand to then calculate the required new inventory that we need to satisfy the expected demand, taking current stock levels into account. The future demand is a random variable that follows a specific probability distribution function. In the context of the newsvendor model (Edgeworth, 1888) (see, e.g. Khouja (1999) for a review), the optimal quantile for a known probability density function can be calculated exactly. Crucially, this approach already assumes that we have chosen a specific probability distribution as our demand model. In many more complex settings in practice, determining the optimal quantile directly often proves to be very challenging. Therefore, more tangible key performance indicators (KPIs) focusing on directly measurable business objectives are used to define the business strategy, such as the stock-out rate or the waste rate of perishable goods. Therefore, if we want to study the effect of different choices of a specific quantile of the demand distribution on the operational decisions such as the order quantity as well as the operationally accessible KPIs, we need access to the full probability density distribution of the future predicted demand. Further, if we plan a deeper analysis on the data-generating process leading to the customer demand, we need access to this variable directly as derived operational quantities such as the optimal order quantity are at best a proxy in the sense of a causal analysis. A wide range of approaches have been suggested that can be broadly categorized in the following way: In the "data-driven newsvendor", the operational quantity is estimated directly from the observable data using, for example, machine learning. In this approach, the demand estimation is no longer modelled explicitly but contained implicitly within the model that is used to estimate the optimal order quantity for a given product at a specific sales location. Although this integrated approach seems at first preferable because it avoids modelling the demand explicitly, this also implies that we no longer have access to the probability distribution for the demand.

*Corresponding author email address: ulrich.kerzel@iu.org
DOI: 10.22034/ijsom.2022.109044.2145

The methods that model demand explicitly can be categorized in model-free and model based approaches. In model-free approaches, we do not assume a specific parameterization or theoretical derivation of the underlying demand model, but estimate this directly from the data. Typical examples of these approaches are ARIMA, LSTM or smoothing techniques that mainly rely on the auto-correlation between observed events to predict future behavior.

Finally, model-based approaches use either a theoretical or empirical model to describe demand as a random variable. This has the advantage that we can base the demand estimate on theoretical considerations as well as obtain a full probability distribution instead of point estimates. However, if we rely on theoretical arguments to motivate our choice of demand model, we need to verify the assumptions of this model and determine if the data support these.

The remainder of the paper is organised as follows:

Sec. 2 summarizes the current status of the relevant literature, sec. 3 discusses the relationship between count data and inter-arrival time, i.e. the way the number of observed sales are linked to the distribution of time between subsequent purchases. Sec. 4 introduces the dataset used in the analysis described in detail in sec. 5, followed by a discussion in sec. 6. The paper ends with a conclusion in sec 7.

1.1. Summary of contributions and goal of the paper

In a model-based approach we need to choose a specific model to forecast future customer demand. A common choice to model customer demand is the negative binomial distribution, also known as a Gamma-Poisson distribution as suggested in Ehrenberg (1959); Goodhardt and Ehrenberg (1967); Chatfield and Goodhardt (1973); Schmittlein, Bemmaor, and Morrison (1985). Implicitly, this assumes that we can treat the sales of goods, for example in a supermarket, as a Poisson process, which in turn implies that the inter-arrival time between subsequent events follows an exponential distribution. Using a public dataset from a supermarket chain, we can investigate if this assumption is supported by data in this scenario. Since the data used in this analysis are recorded per product group and per household, we can investigate the effect of customer heterogeneity on how the purchasing behaviour, and therefore the validity of the model assumptions, varies between different types of products. This topic is not widely explored in the literature, in particular for fast moving goods. However, for the case of slow-moving goods, the nature of the Poisson - process is explored for example for spare part sales in Syntetos, Babai, and Altay (2012); Lengu, Syntetos, and Babai (2014); Turrini and Meissner (2019).

2. Literature Review

The following section summaries the relevant literature and has been grouped into separate sections to make it easier to find the various aspects.

2.1. *Integrated Approaches*

From an operational point of view, the quantity the retailers need to act upon is the optimal order quantity that is transmitted to the manufacturer or wholesaler to procure new goods for future selling periods. Instead of first estimating the future demand and then derive the optimal order quantity, an integrated approach aims at estimating the operational quantity directly. This approach is known as the "data-driven newsvendor" in the recent literature, see e.g. Beutel and Minner (2012); Ban and Rudin (2019); Bertsimas and Kallus (2020); Oroojlooyjadid, Snyder, and Takáĉ (2020); Huber, Müller, Fleischmann, and Stuckenschmidt (2019). At first glance, this approach seems preferable since we no longer need to estimate the probability density function of the future demand as a random variable but can directly estimate the operational quantity from the underlying data, such as, e.g., past sales records, pricing information, promotion campaign, and others. However, the approach that separates the steps of estimating demand first and then derive the optimal order quantity has significant benefits compared to such an integrated approach. If we have access to the full probability distribution describing the predicted future demand, we can analyse the impact of choosing different quantiles as the best point estimators for our ordering decision on operational KPIs. Since such KPIs, for example, waste and stock-out rate in case of perishable food, conflict with each other, analysing the full spectrum of the predicted distribution allows mapping the available choices to the overall business strategy. For example, if the management were to decide that the stock-out rate must not fall below a certain threshold for a specific product group, we can then investigate which quantile we would have to choose to achieve this objective and what the consequences are on, for example, the waste rate. Using the order quantity instead implies that we convolve information about operational aspects beyond our (immediate) control such as lot-sizes and delivery schedules without being able to analyse separately how large the contributions from demand and operational constraints are. Furthermore, from a practitioners perspective, calculating the demand separately from the orders has the benefit that longer-term forecasts can be calculated that can be both shared with other business units as well as used as a failure - recovery mechanism if the forecasting step fails but new orders need to be calculated based on current stock levels and delivery constraints. More importantly, however, the integrated approach does not allow to study the data-generating process, i.e., the mechanism that drives the stochastic behaviour of customer demand. This is paramount if, possibly at a later stage, a causal analysis is planned to study the influence of interventions such as pricing, advertisement or promotion campaigns, or others in either Pearl's do-calculus (Pearl, 2009) or Rubin's potential outcomes framework (Rubin, 1974). This is different from studying such effects in terms of Granger causality (Granger, 1969), which seeks to analyse the effect of one time series on another. While we can, for example, represent the price of a product as a time-series, fundamentally, this is an intervention as retailers actively define a specific price. Such an analysis is only possible if we have access to the quantity directly affected by these interventions, i.e. the predicted demand, as operational quantities such as the optimal order size are, at best, a proxy and do not allow a thorough understanding of the underlying causal relationships and likely lead to unnecessary causal pathways that we may not be able to control for.

2.2. Timeseries Based Approaches

A popular method to estimate future demand is to interpret past sales records, possibly corrected for censored data, as a time-series from which the future demand can be estimated. Generally, this method works by building a regression model from past observations, i.e. $y_{t+1} = f(y_t, y_{t-1}, ...)$. Depending on the approach taken, dedicated components can be introduced to model trends, seasonality, or external factors in a variety of ways. Fundamentally, these approaches exploit the auto-correlation between the variable Y representing the demand at various times t. In this family of approaches, demand is estimated, for example, using auto-regressive integrated moving average (ARIMA) (Box, Jenkins, Reinsel, & Ljungl, 2015) or exponential smoothing methods (Croston, 1972; Holt, 1957; Brown, 1963; Gardner, 1985). A comprehensive review can be found in De Gooijer and Hyndman (2006). Within the context of demand forecasting, examples of these methods can be found in e.g. Huber, Gossmann, and Stuckenschmidt (2017); Kalchschmidt, Verganti, and Zotteri (2006); Fattah, Ezzine, Aman, El Moussami, and Lachhab (2018); Permatasari, Sutopo, and Hisjam (2018). Other approaches are based on Kalman filters (Kalman, 1960; G. W. Morrison & Pike, 1977; Mitropoulos, Samouilidis, & Protonotarios, 1980; Tegene, 1991; Kandananond, 2014; Jacobi, Karimanzira, & Ament, 2007). Building on this concept, state space-models (R. Hyndman, Koehler, Ord, & Snyder, 2008) relate the observable to the evolution of (unknown) states that describe the system of interest, see e.g. Ramos, Santos, and Rebelo (2015) for a comparison of ARIMA vs state-space models within a retail context. Structural time-series (Harvey & Peters, 1990; Taylor & Letham, 2018) models take a similar approach and decompose the observations into trend, seasonality, and further aspects.

Using supervised machine learning, neural networks have been used for time-series forecasting for a long time (Zhang, 2012; Remus & O'Connor, 2001). Machine learning approaches naturally allow to take a wide range of covariates or feature variables into account and methods range from regression tress (Breiman, Friedman, Stone, & Olshen, 1984), to recurrent neural networks (Rumelhart, Hinton, & Williams, 1986), "long short-term memory" (LSTM) networks (Hochreiter & Schmidhuber, 1997), or transformers (Vaswani et al., 2017). The interest in deep learning techniques has significantly increased in recent years within the context of demand forecasting, see e.g. Bandara et al. (2019); Yu, Wang, Strandhagen, and Wang (2017); Goyal, Kumar, Kulkarni, Krishnamurthy, and Vartak (2018); Helmini, Jihan, Jayasinghe, and Perera (2019); Golabek, Senge, and Neumann (2020). A different approach to modelling demand is taken by Alwan, Xu, Yao, and Yue (2016) where the authors use an empirical AR(1) time-series approach.

However, all these methods implicitly or explicitly rely on the auto-correlation between observations of the variable of interest at different times t_i. In many cases, this may be sufficient, however, since these methods rely on learning or exploiting the auto-correlation, they are prone to temporal confounding which may impact both the performance of the model as well as its interpretability. Temporal confounding can arise if some external variable X affects the variable we wish to forecast, for example, the future demand, at different times t_1 and t as illustrated in Fig. 1. If we look at the causal dependency in this directed acyclic graph, the variable X takes the role of a (temporal) confounder, creating a backdoor path between the variable Y at times t_1 and t. Therefore, if we do not control for such a temporal confounder, the auto-correlation of the variable Y between times t_1 and t will at least contain a spurious component introduced by the backdoor path, even if there is a genuine auto-correlation. Neither traditional time-series methods such as ARIMA nor smoothing methods, nor recurrent neural networks or modern LSTM-based can disentangle these effects as they rely on the assumption that the correlations found in the data are genuine.

The methods discussed so far calculate a point estimate, meaning that they forecast a single number as the prediction. While business operations can be based on a single number, this does not allow to estimate the uncertainty on this number. Since the implicit model parameters of, for example, an ARIMA or machine learning model need to be estimated from the available data, these models are intrinsically associated with an uncertainty. A range of approaches has been developed, see e.g. Chatfield (1993) for a review of the methods available at the time. In particular, bootstrapping methods have been studied in detail for time-series forecasting (Masarotto, 1990; McCullough, 1994, 1996; Grigoletto, 1998; Thombs & Schucany, 1990; Clements & Taylor, 2001; Angus, 1994; Pascual, Romo, & Ruiz, 2004, 2001, 2005). However, as already pointed out in Chatfield (1993), estimating the uncertainty is only one aspect of why generally the full probability distribution function of the variable of interest, in our case, the future customer demand, is required because only then can we evaluate the outcome of alternative scenarios described by the quantiles of this distribution on the business strategy. In particular, if we do not wish to make the assumption that the demand follows a Gaussian distribution we either need to assume a specific functional form or estimate the complete distribution directly.

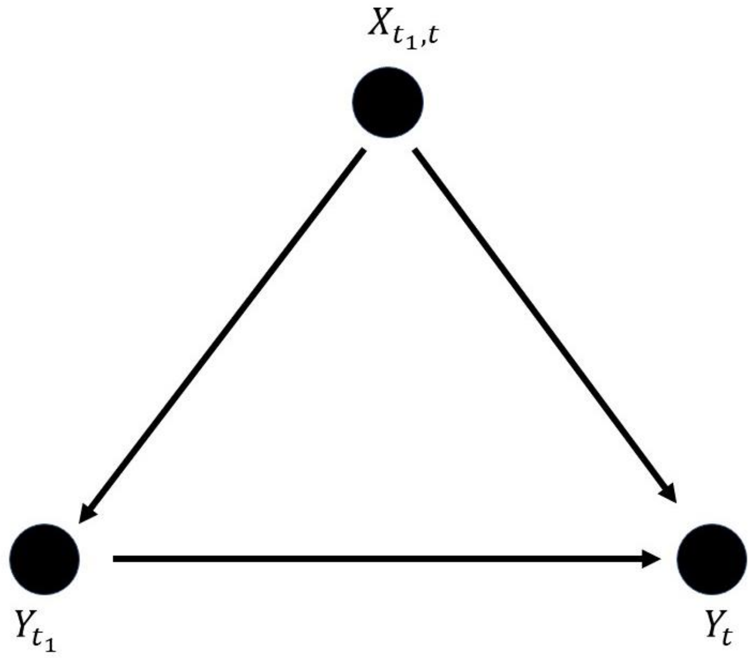


Figure 1: Origin of temporal confounding: An external variable X affects the prediction variable Y at different times t_1 and t, creating a backdoor path between y_{t1} and y_t. The auto-correlation from Y_{t1} to Y_t is therefore a mixture of the true causal effect from Y_{t1} to Y_t as well as the spurious correlation via the backdoor path Y_{t1} via $X_{t1,t}$ to Y_t.

2.3. Model-Free Approaches

In model-free approaches, we generally do not assume that the variable of interest such as the future customer demand can be described by a specific functional form but estimate this from data. This approach is also called quantile regression (Koenker & Hallock, 2001) and can be implemented in several ways. For time-series in particular, a simulation or bootstrapping framework has been developed (R. J. Hyndman, 1995). Methods based on neural networks have also been explored for this purpose (Feindt & Kerzel, 2006; Wen, Torkkola, Narayanaswamy, & Madeka, 2017; Lim, Arik, Loeff, & Pfister, 2019).

2.4. Model-Based Approaches

In model-based approaches, we first choose the underlying probability distribution function that we use to model the variable of interest. Then, for each forecasting step, we need to estimate the model parameters, possibly with their associated uncertainties. Empirically, we could estimate the best fitting function at each forecast step given the current data (Adan, van Eenige, & Resing, 1995), however, due to the stochastic nature of demand, such an approach is not expected to be stable as each new data record may result in a different "optimal" distribution function.

It is therefore advisable to base the choice of probability distribution function on theoretical arguments and then use suitable algorithms such as e.g. Salinas, Flunkert, Gasthaus, and Januschowski (2020) for the modelling. However, each choice for an underlying model is associated with a specific set of assumptions. Even if these models work well in day-to-day operations, it is crucial to verify these assumptions to determine if the assumptions on which the model is built are justified by the observable data.

The count distribution of product sales is typically modelled as a negative binomial distribution (NBD), also known as a Gamma-Poisson distribution. This distribution arises if the parameter λ of is not fixed but a random variable itself that follows a Gamma distribution, i.e. the NBD can be seen of a mixture of Poisson distributions. Compared to the Poisson distribution, the NBD is over-dispersed, i.e. shows a larger variance ($\sigma^2 > \mu$) than the Poisson distribution for which $\mu = \sigma^2$. The model was first developed by Ehrenberg (1959) and discussed further subsequently by e.g. Goodhardt and

Ehrenberg (1967); Chatfield and Goodhardt (1973). Schmittlein et al. (1985) point out why this approach generally works for consumer purchases: Each consumer makes their purchase according to a Poisson process with rate λ and the rates are specific to the individual consumers and follow a Gamma distribution. D. G. Morrison (1969) allows for a large fraction of non-buyers where the peak at zero is inflated beyond the value by a Poisson or NBD model.

Ehrenberg (1959) lists two assumptions: The successive purchase periods are both of equal time and similar to each other, as well as not *too short* such that purchases in one period do not affect the ones in the next period. However, as Chatfield and Goodhardt (1973) points out, a pure Poisson process leads to an exponential distribution as inter-arrival time which implies that the next likely purchase time is immediate, whereas in practice one would expect a dead-time of no purchase after a purchase has been made. This is related to the second assumption by Ehrenberg that the intervals considered should not be too short. In practical situations, *too short* is not very specific and depends on the type of product considered. Alternatively, compound Poisson distributions are used, see e.g. Prak et al. (2018) and references therein.

2.5. Customer Heterogeneity

Customer heterogeneity is one of the key drivers that can lead to large demand fluctuations (Bartezzaghi, Verganti, & Zotteri, 1999) as well as demand patterns that are specific to a particular market. This can lead to an effect known as "lumpy demand", see e.g. Wemmerĺov (1986); Wemmerlōv and Whybark (1984); Ho (1995); Bartezzaghi and Verganti (1995). In these cases, periods of high and low demand alternate in an irregular fashion resulting in "clusters" of customer demand. This behaviour is observed in a range of application areas from, for example, fashion (Fisher, Hammond, Obermeyer, & Raman, 1994) to manufacturing (Brandolese & Cigolini, 1999).

In Kalchschmidt et al. (2006), for example, the authors analyse the impact on business operations in three case studies from spare parts, retail and fresh food and recommend identifying the root causes of customer heterogeneity, define homogeneous groups and then use specific forecasting approaches within each group.

Unlike many case - studies focusing on specific domains, larger supermarkets, in particular, are expected to show a broad range of customer behaviour since the products offered can range from fresh and convenience food, canned or frozen food with a long shelf-life to electronics, fashion, as well as household items. Therefore, this broad range of products sold in supermarkets will likely lead to an equally large range of customer behaviour, since many heterogeneous market domains are combined in the same physical or virtual store. Furthermore, in particular for the case of a physical store, the customer behaviour that would otherwise be different across the broad range of product categories, can overlap since customers may seize the opportunity to buy all items together as a "one-stop-shop". From an operational point of view, supermarket managers in particular are interested in as few operational systems as possible to keep the day-to-day complexity low as reasonably possible. In particular, if we want to base the demand forecast on a single model covering most if not all product lines, we need to investigate whether the assumptions this model makes can be verified in the data.

3. Count Data and Inter-arrival Time

Using a specific model grounded in theoretical arguments is often favourable over model-free, purely data-driven methods to forecast customer demand or integrated approaches for inventory control as discussed in sec. 2. In particular, customer demand is typically modelled as a negative binomial distribution (NBD). The negative binomial model arises if the parameter λ of a Poisson distribution itself is a random variable that is distributed according to a Gamma distribution:

$$\mu \sim \text{Gamma}\left(r, \frac{p}{1-p}\right) \tag{1}$$

where *r* is a form parameter and the rate $\frac{p}{1-p}$ is chosen from the binomial distribution, indicating that this prior describes the total count of $r - 1$ in $\frac{p}{1-p}$ prior observations (Gelman et al., 2013). The Gamma-Poisson model is then the convolution:

$$P(k; r, p) = \int_0^\infty f_{\text{Poisson(k;}\mu)} \cdot f_{\text{Gamma}\left(r, \frac{p}{1-p}\right)} d\mu \tag{2}$$

Inserting the definition of the Poisson and Gamma distribution into Eq. (2), we obtain:

$$\begin{aligned}
P(k;r,p) &= \int_0^\infty f_{\text{Poisson(k;}\mu)} \cdot f_{\text{Gamma}\left(r,\frac{p}{1-p}\right)} d\mu \\
&= \int_0^\infty \left[\frac{\mu^k e^{-\mu}}{k!}\right] \times \\
&\left[\frac{1}{\Gamma(r)\left(\frac{p}{1-p}\right)^r}\mu^{r-1}e^{-\mu(1-p)/p}\right] d\mu \\
&= \frac{(1-p)^r p^{-r}}{k!\Gamma r}\int_0^\infty \mu^{r-1+\mu}e^{-\mu/p}d\mu
\end{aligned}$$

Using the identity

$$\int_0^\infty y^b e^{-ay}dy = \frac{\Gamma(b+1)}{a^{b+1}}$$

we can evaluate the integral

$$\int_0^\infty \mu^{r-1+\mu}e^{-\mu/p}d\mu = p^{r+k}\Gamma(r+k)$$

and with $\Gamma(x + 1) = x!$, we can then write:

$$\begin{aligned}
P(k;r,p) &= \frac{(1-p)^r p^{-r}}{k!\Gamma(r)}p^{r+k}\Gamma(r+k) \\
&= \frac{\Gamma(r+k)}{\Gamma(r)k!}p^k(1-p)^r \\
&= \frac{(r+k-1)!}{(r-1)!k!}p^k(1-p)^r \\
&= \binom{k+r-1}{k}p^k(1-p)^r \\
&= \text{Negative} - \text{Binomial}(r,p)
\end{aligned}$$

In most cases, we take the observed sales data as a proxy for the customer demand where we may have to account for censored data in case of stockout situations. Fundamentally, therefore, the underlying process behind the observed data is a Poisson process. The count distribution of a Poisson or NBD model hence answers the question: *How many products are being sold?* A related question which probes the underlying assumptions of the model is given by *What is the time between purchases?*. For a Poisson-type process, this inter-arrival time is described by a continuous exponential distribution which can be seen as follows: The Poisson distribution is given by:

$$P(r;\mu) = \frac{\mu^r e^{-\mu}}{r!}$$

with mean μ. In a Poisson process, i.e. when the events occur at an average rate λ per time interval, $\lambda \cdot t$ events will have happened during time t. The Poisson distribution for this is then

$$P(r) = (\lambda t)^r e^{-\lambda t}/r!\ .$$

The probability that no events occur, i.e. $r = 0$, is then given by $P(0) = e^{-\lambda t}$. The probability for an event to occur in time t is given by

$$P(T \leq t) = 1 - P(r = 0|\mu = \lambda t) = 1 - e^{-\lambda t}$$

which is the cumulative density distribution of the exponential distribution, i.e. the corresponding probability density distribution is $f(t) = \lambda e^{-\lambda t}$. Hence the Poisson-type distributions and the exponential distributions are linked in the following way: If the count data of the events can be described by a Poisson-type distribution, the time between two events follows an exponential distribution. If we focus instead of the next purchase on the interval after which multiple purchases have occurred, the process can be described by an Erlang distribution as discussed in e.g. Chatfield and Goodhardt (1973).

4. Data

The data are taken from a public dataset (Venkatesan, 2014). They consist of anonymous transactions at a supermarket chain at the level of households and have been collected over a period over two years. In total, data from approximately 2,500 households were recorded who were frequent shoppers at this retailer. The retailer at which the data were taken is not disclosed in this public dataset. However, as product sizes for liquids are given in ounces (OZ) and the data include a product hierarchy detailed as "Hispanic", the data are likely recorded in the Americas. The data contain all purchases of the households across all product categories and are organised in a range of tables that describe the details of the product, transaction data (when was which product bought by which household), as well as further data tables that contain details to marketing campaigns. In this analysis, the data tables relating to transactional data and product details are relevant. The product details include department, product group name, sub-group name and its size, for example: (department = grocery, product group = household cleaning needs, product sub-group = ammonia) or (department = grocery, product group = fluid milk products, product sub-group = fluid milk white only). This means that while the individual product or brand cannot be specified, the type of item which the consumer bought can be identified to a fine granularity. This implies that we cannot identify aspects such as brand loyalty for individual products and neither the demographics and shopping preferences of customers within a single household.

The table containing these data can be joined with the table holding the transactional data which allows to identify what a member of a specific household has bought when. These transactional data contain an identifier of the household that has bought the relevant product that in turn is linked to the table holding the product information using a product ID, as well as the quantity bought and an identification detailing the purchase time. After the join, the relevant data looks like:

Day	Quantity	Household ID	Department	Product Group	Product Sub-Group
217	*1*	*1675*	*GROCERY*	*HOUSEHOLD CLEANG NEEDS*	*AMMONIA*
46	*1*	*1476*	*GROCERY*	*FLUID MILK PRODUCTS*	*FLUID MILK WHITE ONLY*

A further field indicates the product ID that identifies a specific SKU even though a clear-text description of this product is not provided.

Because the data are recorded at the level of households, we cannot distinguish between the shopping behaviour of, say, children and parents within the same household. We can, however, analyse how the purchasing behaviour of individual household changes across product groups which is sufficient to achieve our objective to determine if the assumptions that underlay the negative binomial model typically used in model-based forecasting approaches are fulfilled across a range of product types with very different characteristics.

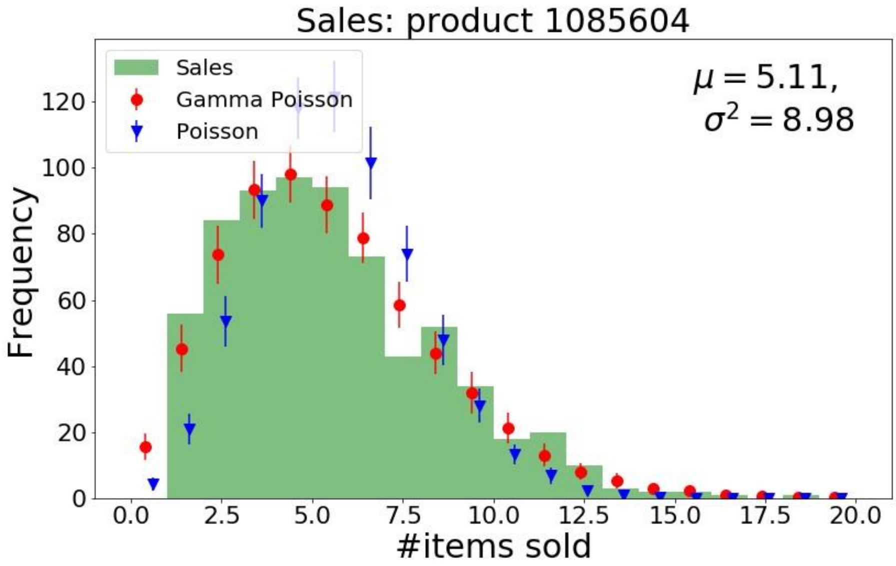


Figure 2. Count data for an individual product well described by both a Poisson and a NBD model.

5. Analysis

Following the discussion in Sec. 2 and 3, we expect that the count data of individual products can be described by a Poisson or more generally by a negative binomial model. Apart from theoretical arguments, we expect that in most cases the NBD model will describe the data better as the negative binomial distribution is over-dispersed, i.e. $\sigma^2 > \mu$ compared to the Poisson distribution, for which $\sigma^2 = \mu$. This means that in practice the additional parameter allows for longer tails. Figs. 2 and 3 show two examples for the overall count data of the sales of individual products. Overlaid on the data (histogram, green) are curves for a Poisson (triangle, blue) and Gamma-Poisson or NBD (circle, red) which have been computed using *scikit-learn* (Pedregosa et al., 2011). The curves were generated in the following way: In the first step, the mean and variance were determined from the data. The mean is used directly for the Poisson model, the parameters p and r for the NBD are calculated according to Eqn. (3).

$$p = \mu/\sigma^2 \text{ and } r = \mu^2/(\sigma^2 - \mu) \quad (3)$$

Since the number of count data are limited, bin-to-bin fluctuations are expected when comparing the data to the models in a fit. In order to estimate the spread of distributions generated using these parameters, 50 toy experiments were performed for both the Poisson and the NBD model with the same parameters, using the total number of observed count data as the sample size in each experiment. Then, the average of these values in each bin was taken to represent the central point in the figure, denoted by a triangle for the Poisson model and a circle for the NBD model. The error bars in each bin reflect the spread ($\sigma = \sqrt{var}$) of the generated distributions. As we can see from Figs. 2 and 3, the NBD is able to describe the observed data quite well, whereas a pure Poisson model may fit in some cases but has typically too narrow tails to capture the observed sales patterns. Since the NBD Model belongs to the family of Poisson models, we expect in general that the inter-arrival time between two subsequent events follows an exponential model, as already discussed in Sec. 3. Looking at the customer behaviour from a bird's eye view across all products, stores and households, Fig. 4 describes this assumption reasonably well.

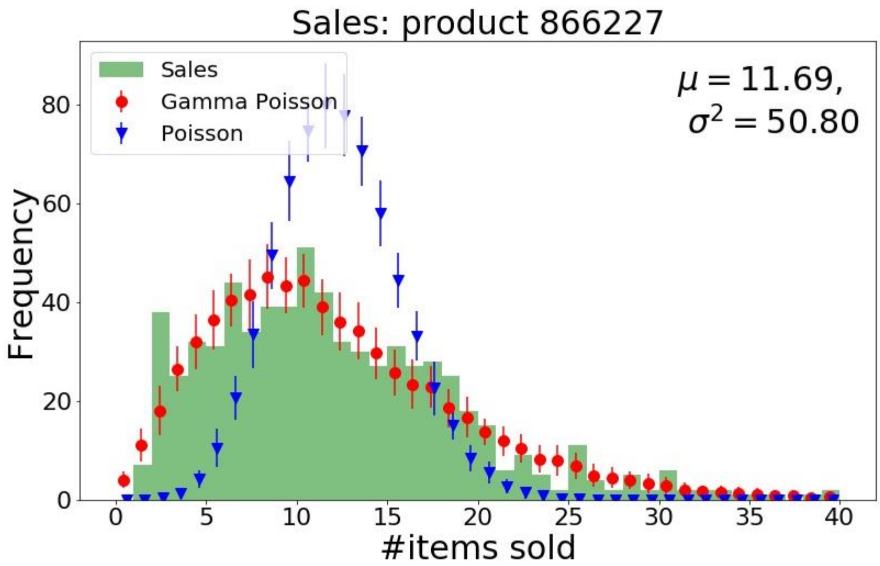


Figure 3. Count data for an individual product well described by a NBD model but overdispersed for a Poisson model.

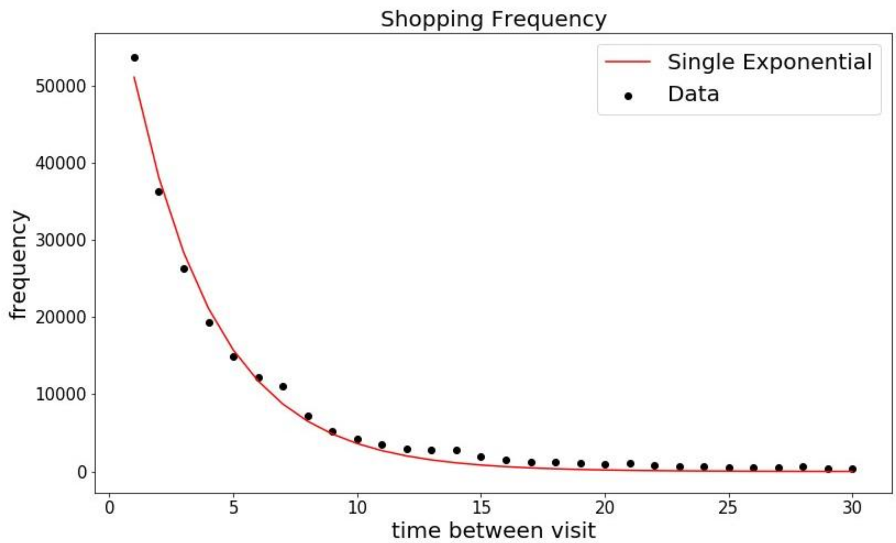


Figure 4. Inter-arrival time of all items bought by individual households. The data are fitted with an exponential curve.

The fit to the function $f(t) = ae^{-\lambda t}$ describes the data well except for two areas at day 7 and around day 14 where the data points deviate from the expected exponential distribution: If 7 or multiples of 7 days have passed between shopping trips, customers are in general slightly more likely to go shopping compared to any other interval. The effect is most pronounced for a 7-day interval and becomes less distinct for 14, 21, ... days. This means that overall the shopping patterns are compatible with the assumption of a Poisson process when we look at the data aggregated across all customers and all products. However, the deviations at day 7 and around day 14 already indicate that the underlying process is not quite as random as assumed in a Poisson-type process but customers have a preferred day on which they do the bulk of their shopping.

The dataset contains shopping records for individual households as well as detailed information about the product hierarchy. While this does not allow to determine which specific product has been bought by a particular individual customer, this data does allow detailed insights into the shopping behaviour related to different types of products. Rather than looking at the bird's eye view from the perspective of the retailer, the data can be analysed from the perspective of single households: How often do members of this household go shopping and how does this depend on the product category? A typical supermarket has a large number of different products (up to tens of thousands for large super- or hypermarkets across a wide range of product categories), leading to a large number of product categories and sub-categories. In the public dataset analysed here, products are grouped into 1000 different sub-categories. In the following, three representative sub-categories have been chosen which each represent very different items bought by a typical household in the area where the supermarkets are located:

- Milk (Fluid milk products: aseptic & milk drinks, fluid milk white only, premium pints)
- Deli & Prepared Food
- Household Cleaning Items

The first category (milk products) is representative of fresh food that has a shelf-life of typically a few days, whereas the category "Deli & Prepared Food" typically consists of convenience food intended for immediate or same-day consumption. The final category (household cleaning items) represents products with a long shelf-life that are typically not prone to impulsive buying patterns.

Whereas Fig. 4 showed that the overall assumption that purchases in a supermarket follow the expected pattern from a Poisson process, we can see from Fig. 5 that the shopping behaviour depends strongly on the product sub-category.

For example, the inter-arrival time of products in the sub-category *Deli & Prepared Food* follow an almost exponential behaviour as expected by the Poisson nature of the shopping process. The 7-day periodicity is visible but not very pronounced. This can be explained by the fact that products in this category are made for (almost) immediate consumption, such as pre-prepared lunch or dinners: Customers buy a ready-made meal and may do so again on the next day.

On the other hand, the purchasing behaviour for household cleaning items is completely different: Since these products last typically a week or more, the corresponding inter-arrival time shows a strong 7-day periodicity where the peaks at 7 and 14 days have almost the same height, falling off gently towards higher periods such as 21 and 28 days. There is no underlying exponential decay in the observed inter-arrival times, i.e. products from these and similar categories are never bought "randomly" as expected by a Poisson-type model but mostly on fixed and regular shopping days.

Most other products and sub-categories are between these extremes. A suitable example is milk products recorded in the data: The time between subsequent purchases show a strong. 7 day period and a weaker 14,21,28 day period owing to the fact that (processed) milk will typically last a few days once opened. Notably, from about 7 days onward the distribution shows an exponential decay in the inter-arrival time indicating a Poisson-type contribution to the shopping pattern. However, the time between purchases in the regime day 1-6 does not follow an exponential function but shows an "activation curve" specific to the types of products sold. In this case, the supermarket likely doesn't stock fresh milk which only lasts for a day or two but only processed milk with a longer shelf life. However, the exact behaviour depends on the specific properties of the individual product.

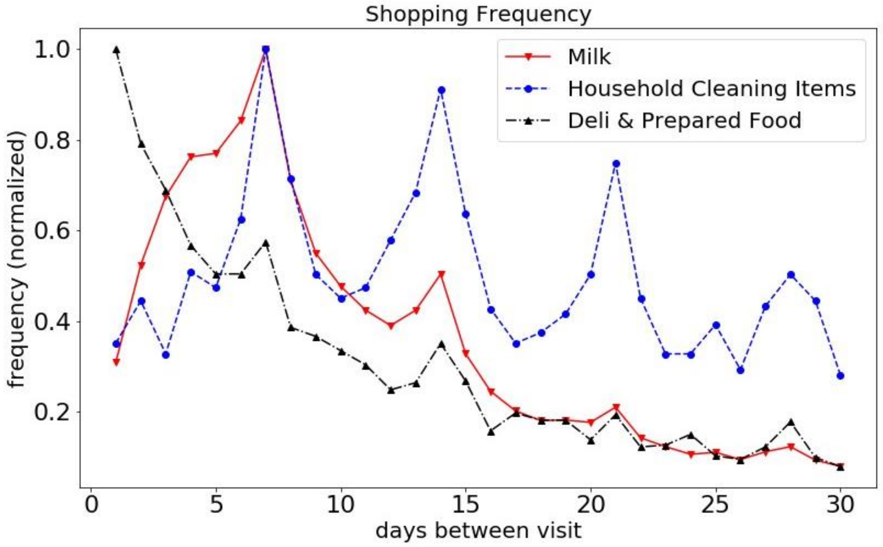


Figure 5. Inter-arrival times for different Product sub-groups for individual households. The lines connect the data points and are added to guide the eye.

6. Discussion

The key decision for any model-based demand forecasting approach is to decide which model should be used to describe the future sales of products. Typically, a Gamma-Poisson or negative binomial distribution is used to express the probability distribution governing the future demand (Ehrenberg, 1959; Goodhardt & Ehrenberg, 1967; Chatfield & Goodhardt, 1973). This assumes that the purchase process as such is a Poisson process (Schmittlein et al., 1985), which in turn implies that the interarrival time between subsequent purchases follows an exponential distribution (Chatfield & Goodhardt, 1973).

When looking at the sales of all products bought by individual households, we can see that this expectation is generally well reproduced in the actual data as shown in Fig. 4. The time between subsequent purchases can generally be described by a single exponential function as expected from the theoretical motivation. However, there are notable exceptions: At day 7, day 14 and, to a lesser extend, day 21 the number of purchases increases and deviates visibly from the expectation of a single exponential function. This means that most households have a fixed "shopping day" at which they so most of their weekly shopping.

However, when looking at the granularity of sub-categories, a more nuanced picture emerges as shown in Fig. 5. The time between purchases of products with a very short shelf-life that perish very quickly follow the expected exponential function. This means that the purchase of these products is dominated by the stochastic aspect of demand and can be interpreted as a pure Poisson process.

The other extreme are products with a very long shelf-life that also last a long time once purchased. The sub-category of household cleaning items can be used to illustrate this purchasing behaviour: The time between purchases does not show the expected exponential function but instead, purchasing events are separated by clearly visible 7-day intervals. This means that these products are bought almost exclusively at regular shopping trips and will last at least until the next planned trip.

Most products are between the extremes of a pure Poisson process and an almost regular pattern which is illustrated by the sub-category of milk: The longer shelf-life allows for a more regular pattern whereas the visible exponential behaviour highlights the presence of the underlying Poisson process. The purchasing behaviour for these products also exhibits the expected dead-time (Chatfield & Goodhardt, 1973), during which no immediate repeat purchase is required.

Overall, the data show that choosing the popular negative binomial model to describe customer demand are well founded in theoretical arguments. In particular, products with a short shelf-life intended for immediate consumption show the exponential behaviour expected from a pure Poisson-type process. However, the longer the shelf-life and the less the products are intended for immediate consumption, the more the the purchase of these products is dominated by a more predictable weekly pattern. Retailers need to take this into account in their demand forecast, for example, by adding a adding a weekly seasonal component that reflects the shopping habits of their customers. This behaviour is similar to the arrival process in queuing theory, combining scheduled arrivals with a varying degree of random arrivals, where the mixture depends on the product type. However, typically queuing theory is used to optimise, for example, the check-out waiting times as in Priyangika and Cooray (2015).

The relation between the count data and arrival times is not widely explored in the literature. Syntetos et al. (2012), Lengu et al. (2014) and Turrini and Meissner (2019) investigate the demand arrival process for slow moving spare parts. These studies also find that a Poisson-type process provides a good description of the data, however, compared to the data in this study, slow moving spare parts are characterised by inter-arrival times of weeks and months rather than days.

7. Conclusion

Demand forecasting continues to be of vital interest for retailers and academic research. Model based approaches can offer substantial benefits compared to model-free or purely data-driven approaches. However, each specific model chosen is associated with a set of underlying assumptions.

Retail demand is typically modelled using a negative binomial distribution which implies that the underlying process is a Poisson-type process. Using a public dataset recording the shopping patterns of individual households in a large supermarket, the underlying assumptions of the Poisson process, in particular, the exponential distribution of the inter-arrival time between subsequent purchasing events has been probed across multiple product categories. Looking at an aggregate level across all product categories, the time between two purchases can be described by an exponential model as expected, with significant deviations at a 7 and 14 day period. This indicates that customers generally have a preferred weekly shopping day at which the bulk of their shopping is done. When looking at product categories with very different characteristics such as convenience food and cleaning items, the data show that the purchasing behaviour is very different across these categories. Whereas convenience food tends to follow the exponential distribution expected from a Poisson - type process, cleaning items are typically bought on regular weekly shopping trips.

This implies that while the negative binomial model can give a good description of customer demand in retail and its underlying assumptions are well or approximately fulfilled for some types of products, this is not necessarily the case across all product types and categories. The shopping behaviour of individual household does not generally follow a Poisson - type process, even if this approach can give good results as an effective model in practice.

The data used in this study are recorded at the level of households and product sub-group. This allows a detailed understanding of the implicit assumptions of the Poisson-type process that is typically assumed in modelling the underlying probability distribution describing the demand. In particular, this allows to understand whether using the commonly used negative binomial model is suitable for modelling the demand of this product category. This granularity of the data is also sufficient to target households with specific marketing approaches at this level. For example, the marketing department could use this information to issue coupons or smartphone notifications pertaining the purchase behaviour of this households for specific product groups. However, a even more detailed customer targeting at the level of individuals and specific products (as opposed to the product sub-category) the granularity of the recorded data would have to be increased. Before implementing a demand forecasting and replenishment solution for a specific retail chain it should be checked if equivalent data for this retailer is available and use this to tailor the models more specifically towards this retailer. If no such data are available, the conclusions from this study can be used but it should be investigated whether obtaining similar data in the future is possible. Retailers will typically have access to the sales data at SKU level but customer tracking at the level of individual purchases are more difficult in most countries. Further avenues of research could investigate who methods from queuing theory can be integrated in demand forecasting models.

References

Adan, I., van Eenige, M., and Resing, J. (1995). Fitting discrete distributions on the first two moments. *Probability in the engineering and informational sciences*, Vol. 9(4), pp. 623–632.

Alwan, L. C., Xu, M., Yao, D., and Yue, X. (2016). The dynamic newsvendor model with correlated demand. *Decision Sciences*, Vol. 47(1), pp. 11-30.

Angus, J. E. (1994). The probability integral transform and related results. *SIAM Review*, *36*(4), pp. 652–654.

Ban, G.-Y., and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, Vol. 67(1), pp. 90–108.

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., and Seaman, B. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In *International conference on neural information processing* (pp. 462–474).

Bartezzaghi, E., and Verganti, R. (1995). Managing demand uncertainty through order overplanning. *International Journal of Production Economics*, Vol. 40(2-3), pp. 107–120.

Bartezzaghi, E., Verganti, R., and Zotteri, G. (1999). A simulation framework for forecasting uncertain lumpy demand. *International Journal of Production Economics*, Vol. 59(1-3), pp. 499– 510.

Bertsimas, D., and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, Vol. 66(3), pp. 1025–1044.

Beutel, A.-L., and Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, Vol. 140(2), pp. 637–645.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljungl, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Brandolese, A., and Cigolini, R. (1999). A new model for the strategic management of inventories subject to peaks in market demand. *International Journal of Production Research*, Vol. 37(8), pp. 1859–1880.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Brown, R. G. (1963). *Smoothing forecasting and prediction of discrete time series*. Prentice-Hall International, Inc. London, UK.

Chatfield, C. (1993). Calculating interval forecasts. *Journal of Business & Economic Statistics*, 11(2), pp. 121–135.

Chatfield, C., and Goodhardt, G. J. (1973). A consumer purchasing model with erlang interpurchase time. *Journal of the American Statistical Association*, *68*(344), pp. 828-835.

Clements, M. P., and Taylor, N. (2001). Bootstrapping prediction intervals for autoregressive models. *International Journal of Forecasting*, Vol. 17(2), pp. 247–267.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, Vol. 23(3), pp. 289–303.

De Gooijer, J. G., and Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, Vol. 22(3), pp. 443–473.

Edgeworth, F. (1888). The mathematical theory of banking. *Journal of the Royal Statistical Society*, Vol. 53, pp. 113-127.

Ehrenberg, A.S., 1959. The pattern of consumer purchases. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 8(1), pp.26-41.

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., and Lachhab, A. (2018). Forecasting of demand using arima model. *International Journal of Engineering Business Management*, Vol. 10, pp. 1-9.

Feindt, M., and Kerzel, U. (2006). The neurobayes neural network package. *NIM A*, Vol. 559(1), pp. 190–194.

Fisher, M. L., Hammond, J. H., Obermeyer, W. R., and Raman, A. (1994). Making supply meet demand in an uncertain world. *Harvard business review*, Vol. 72, pp. 83–83.

Gardner, E. S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*, Vol. 4, pp. 1–28.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Golabek, M., Senge, R., and Neumann, R. (2020). Demand forecasting using long short-term memory neural networks. Preprint on ArXiv. arXiv: 2008.08522

Goodhardt, G. J., and Ehrenberg, A. (1967, May). Conditional trend analysis: A breakdown by initial purchasing level. *Journal of Marketing Research*, Vol. *IV*, pp. 155-161.

Goyal, A., Kumar, R., Kulkarni, S., Krishnamurthy, S., and Vartak, M. (2018). A solution to forecast demand using long short-term memory recurrent neural networks for time series forecasting. In *Midwest decision sciences institute conference*. pp. 12-14.

Granger, C. W. J. (1969, August). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, Vol. 37(3), pp. 424.

Grigoletto, M. (1998). Bootstrap prediction intervals for autoregressions: some alternatives. *International Journal of Forecasting*, Vol. 14(4), pp. 447–456.

Harvey, A., and Peters, S. (1990). Estimation procedures for structural time series models. *Journal of Forecasting*, Vol. 9, pp. 89–108.

Helmini, S., Jihan, N., Jayasinghe, M., and Perera, S. (2019). Sales forecasting using multivariate long short term memory network models. *PeerJ PrePrints*, Vol. 7, pp. e27712v1.

Ho, C.-J. (1995). Examining the impact of demand lumpiness on the lot-sizing performance in mrp systems. *International Journal of Production Research*, Vol. 33(9), pp. 2579–2599.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, Vol. 9(8), pp. 1735–1780.

Holt, C. C. (1957). Forecasting trends and seasonal by exponentially weighted moving averages. *ONR Memorandum*, Vol. 52, pp. 5-10.

Huber, J., Gossmann, A., and Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert systems with applications*, Vol. 76, pp. 140–151.

Huber, J., Mu¨ller, S., Fleischmann, M., and Stuckenschmidt, H. (2019). A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research*, Vol. 278(3), pp. 904–915.

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

Hyndman, R. J. (1995). Highest-density forecast regions for nonlinear and non-normal time series models. *Journal of Forecasting*, Vol. 14(5), pp. 431–441.

Jacobi, M., Karimanzira, D., and Ament, C. (2007). Water demand forecasting using Kalman filtering. In *Proceedings of the 16th iasted international conference on applied simulation and modelling* (pp. 199–202).

Kalchschmidt, M., Verganti, R., and Zotteri, G. (2006). Forecasting demand from heterogeneous customers. *International Journal of operations & Production management*. Vol. 26 (6), pp. 619-638.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, Vol. 82(1), pp. 35-45.

Kandananond, K. (2014). Applying Kalman filter for correlated demand forecasting. *Applied mechanics and materials* Vol. 619, pp. 381–384.

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

Koenker, R., and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, Vol. 15(4), pp. 143–156.

Lengu, D., Syntetos, A. A., and Babai, M. Z. (2014). Spare parts management: Linking distributional assumptions to demand classification. *European Journal of Operational Research*, Vol. 235(3), pp. 624–635.

Lim, B., Arik, S. O., Loeff, N., and Pfister, T. (2019). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of forecasting*, Vol. 37(4), pp. 1748-1764.

Masarotto, G. (1990). Bootstrap prediction intervals for autoregressions. *International Journal of Forecasting*, Vol. 6(2), pp. 229–239.

McCullough, B. (1994). Bootstrapping forecast intervals: an application to AR(p) models. *Journal of Forecasting*, Vol. 13(1), pp. 51–66.

McCullough, B. (1996). Consistent forecast intervals when the forecast-period exogenous variables are stochastic. *Journal of Forecasting*, Vol. 15(4), pp. 293–304.

Mitropoulos, C., Samouilidis, J., and Protonotarios, E. (1980). Using Kalman filtering for energy forecasting. *IFAC Proceedings Volumes*, Vol. 13(5), pp. 317–324.

Morrison, D. G. (1969). Conditional trend analysis: A model that allows for nonusers. *Journal of Marketing Research*, Vol. 6(3), pp. 342-346.

Morrison, G. W., and Pike, D. H. (1977). Kalman filtering applied to statistical forecasting. *Management Science*, Vol. 23(7), pp. 768–774.

Oroojlooyjadid, A., Snyder, L. V., and Taka´˘c, M. (2020). Applying deep learning to the newsvendor problem. *IISE Transactions*, Vol. 52(4), pp. 444–463.

Pascual, L., Romo, J., and Ruiz, E. (2001). Effects of parameter estimation on prediction densities: a bootstrap approach. *International Journal of Forecasting*, Vol. 17(1), pp. 83–103.

Pascual, L., Romo, J., and Ruiz, E. (2004). Bootstrap predictive inference for arima processes. *Journal of Time Series Analysis*, Vol. 25(4), pp. 449–465.

Pascual, L., Romo, J., and Ruiz, E. (2005). Bootstrap prediction intervals for power-transformed time series. *International Journal of Forecasting*, Vol. 21(2), pp. 219–235.

Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge University Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.

Permatasari, C. I., Sutopo, W., and Hisjam, M. (2018). Sales forecasting newspaper with arima: A case study. In *Aip conference proceeding*, Vol. 1931, p. 030017.

Prak, D., Teunter, R. H., Babai, M. Z., Syntetos, A. A., Boylan, J.E. (2018). Forecasting and inventory control with compound poisson demand using periodic demand data. *SOM Research Reports*, Vol. 10.

Priyangika, J., & Cooray, T. (2015). Analysis of the sales checkout operation in supermarket using queuing theory. *Universal journal of management*, Vol. 4(7), pp. 393-396.

Ramos, P., Santos, N., and Rebelo, R. (2015). Performance of state space and arima models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing*, Vol. 34, pp. 151–163.

Remus, W., and O'Connor, M. (2001). Neural networks for time-series forecasting. In *Principles of forecasting* (pp. 245–256). Springer.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, Vol. 66(5), pp. 688–701.

Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, Vol. 323, pp. 533–536.

Salinas, D., Flunkert, V., Gasthaus, J., and Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, Vol. 36(3), pp. 1181–1191.

Schmittlein, D. C., Bemmaor, A. C., and Morrison, D. G. (1985). Technical note—why does the nbd model work? robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Science*, Vol. 4(3), pp. 255-266.

Syntetos, A. A., Babai, M. Z., and Altay, N. (2012). On the demand distributions of spare parts. *International Journal of Production Research*, Vol. 50(8), pp. 2101–2117.

Taylor, S. J., and Letham, B. (2018). Forecasting at scale. *The American Statistician*, Vol. 72(1), pp. 37–45.

Tegene, A. (1991). Kalman filter and the demand for cigarettes. *Applied Economics*, Vol. 23(7), pp. 1175–1182.

Thombs, L. A., and Schucany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, Vol. 85(410), pp. 486–492.

Turrini, L., and Meissner, J. (2019). Spare parts inventory management: New evidence from distribution fitting. *European Journal of Operational Research*, Vol. 273(1), pp. 118–130.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., and Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), *Advances in neural information processing systems,* Vol. 30 (pp. 5998–6008). Curran Associates, Inc.

Venkatesan, R. (2014). *The complete journey*. https://www.dunnhumby.com/careers/ engineering/sourcefiles. (Visited on 2019-Sep-10)

WemmerÍov, U. (1986). A time-phased order-point system in environments with and without demand uncertainty: a comparative analysis of non-monetary performance variables. *International journal of production research*, Vol. 24(2), pp. 343–358.

WemmerÍov, U., and Whybark, D. C. (1984). Lot-sizing under uncertainty in a rolling schedule environment. *The International Journal Of Production Research*, Vol. 22(3), pp. 467–484.

Wen, R., Torkkola, K., Narayanaswamy, B., and Madeka, D. (2017). A multi-horizon quantile recurrent forecaster. *31st conference on Neural information processing systems (NIPS) arXv:1711-11053.*

Yu, Q., Wang, K., Strandhagen, J. O., and Wang, Y. (2017). Application of long short-term memory neural network to sales forecasting in retail—a case study. In *International workshop of advanced manufacturing and automation* (pp. 11–17).

Zhang, G. (2012). Neural networks for time-series forecasting. In *Handbook of natural computing* (pp. 461–477). Springer Berlin, Heidelberg.