European Journal of Operational Research 303 (2022) 255–272

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Production, Manufacturing, Transportation and Logistics

An empirical Bayes approach to incorporating demand intermittency and irregularity into inventory control

Yuan Ye[a,*], Yonggang Lu[b], Powell Robinson[c], Arunachalam Narayanan[d]

[a] Assistant Professor, College of Business, Department of Marketing and Supply Chain Management, California State University, Sacramento
[b] Associate Professor, Maine Business School, University of Maine
[c] Professor, C.T. Bauer College of Business, Department of Decision and Information Sciences, University of Houston
[d] Associate Professor, G Brint Ryan College of Business, Department of Information Technology and Decision Sciences, University of North Texas

ARTICLE INFO

Article history:
Received 3 February 2021
Accepted 16 February 2022
Available online 22 February 2022

Keywords:
Forecasting
Inventory management
Spare parts
Empirical Bayes

ABSTRACT

Spare parts inventory management is complex due to the combined impact of intermittent and variable demand patterns. It becomes even more challenging if the spare parts demand distribution is highly complex due to strong interdependent demand intermittency and extremely irregular demand. The research literature proposes many analytical methods for forecasting spare parts demand. But, due to their limited flexibility in modeling complex demand patterns, existing forecasting methods may not produce satisfactory results for a spare parts portfolio displaying extremely complex demand patterns. This study proposes a novel nonparametric Bayesian forecasting approach with its roots in the empirical Bayes paradigm. The method is subject to few performance constraints and is highly flexible in dealing with a rich diversity of demand patterns, including extreme demand complexity. We assess the relative performance of this new approach with several prominent methods in the literature using an automotive parts distributor's empirical demand data for 46,272 stock-keeping units. This dataset is representative of typical spare parts portfolios that are characterized by a wide variety and extremely complex demand patterns. The experimental findings show the new Bayesian approach achieves the best overall performance in terms of inventory efficiency and minimal backorders for meeting specified target service levels. This favorable performance reflects the approach's flexibility to accommodate disparate and complex demand patterns, including interdependence of demand intermittency, irregular demand distribution, and even nonstationary demand distribution to some extent, and provide robust solutions.

1. Introduction

Effectively managing spare parts inventory for product and equipment maintenance is crucial for many manufacturing and service organizations, such as those in automotive, industrial products, telecommunications, medical devices, and transportation industries among others. As noted in Deloitte Consulting (2013), the spare parts business is the main driver to enhance customer satisfaction and generate repurchase opportunities in many firms. Estimates of the global spare parts market size range from $700 billion to $1.5 trillion (Desomer, 2011; Jasper, 2006). Because of relatively high profit margins, a company's spare parts business unit alone may account for up to 75 percent of the firm's overall profits (Desomer, 2011). Thus, properly managing spare parts inventory is essential to ensure smooth operations and boost the firm's bottom-line.

However, intermittent demand patterns; i.e., random demand with a large portion of zero values and highly erratic nonzero demand, are common among spare parts, which complicates forecasting (Boylan & Syntetos, 2010; Cattani, Jacobs & Schoenfelder, 2011) and ultimately inventory management (Bacchetti & Saccani, 2012; Wouters & Rustenburg, 2014). In a Delphi study of senior service parts managers, Boone, Craighead and Hanna (2008) identify inaccurate forecasts as one of the top two challenges confronting service parts inventory management.

Often, the literature considers demand forecasting and inventory management as separate problems; that is, forecasting literature ignores the impact of the forecast method on inventory levels and customer service, while the inventory literature assumes the demand distribution and its parameters are known (Goltsos, Syntetos, Glock & Ioannou, 2021; Syntetos, Babai & Gardner, 2015). However, accounting for the interactions between the forecasting

* Corresponding author.
E-mail address: yuan.ye@csus.edu (Y. Ye).

https://doi.org/10.1016/j.ejor.2022.02.033

method and inventory system is imperative for efficient inventory management.

Consider a multi-period inventory management system for a spare part with intermittent demand and nonzero replenishment lead time. Because of high demand uncertainty, safety stock inventory is held to protect against shortages during the replenishment cycle. The objective is to determine the minimum inventory position (on-hand and on-order inventory) at order placement to ensure a specified service level as defined by either the probability of not stocking out during the replenishment cycle or the fraction of demand supplied from on-hand inventory. This decision requires predicting the replenishment cycle demand (RCD) distribution during which time the inventory system is vulnerable to stock out. The replenishment cycle length for the continuous review, fixed order quantity inventory model (Q, r) is the replenishment lead time (L); and for the periodic review (T, R) model with order-up-to level R and review interval (T), the replenishment cycle length is $L+\mathrm{T}$. Regardless of the inventory control system, the appropriate choice of forecast method and parameters can reduce the replenishment cycle's forecast error and consequent safety stock (Boylan & Syntetos, 2010).

The two main schools of thought for intermittent demand forecasting embrace either a parametric or nonparametric approach. Parametric methods, such as simple exponential smoothing (SES), Croston's (1972) method (CM) and some recent Bayesian methods, assume a known standard demand distribution (e.g., Normal, Poisson, compound Poisson, uniform) per unit time and estimate its parameters, typically the mean and variance, to predict the replenishment cycle's demand. Nonparametric methods do not rely on any demand distributional assumption. For instance, nonparametric bootstrapping methods randomly resample historical demand data and construct an empirical frequency distribution of replenishment cycle demand. The nonparametric domain also includes neural network (NN) methods, which are data driven and free of restrictive probabilistic model assumptions (Babai, Tsadiras & Papadopoulos, 2020; Boylan & Syntetos, 2021). There is no consensus in the literature on whether the parametric or nonparametric approach is best suited for all industry applications. Instead, the relative performance of the methods appears to be determined by how well the model's features (particularly, its demand distribution assumptions) fit the particular characteristics of industrial problems.

This study is motivated by research collaboration with a heavy truck spare parts distributor. The distributor seeks to maintain a high spare parts in-stock inventory service level at minimal inventory investment at their distribution centers. Achieving high forecast accuracy at the item level plays a key role in balancing these tradeoffs. In addition to serving as industry knowledge experts, the firm assembled a dataset consisting of three years of bi-weekly demand data for 46,272 SKUs stocked at the firm's largest distribution center. This empirical dataset is distinguished from others used in spare parts forecasting research due to the wide variety and complexity of the SKU's demand patterns, which range from no demand intermittency to many consecutive periods of demand intermittency and from standard, stationary nonzero demand distributions to highly irregular and/or nonstationary demand distributions. As a result, the distributor is unable to achieve its goals using currently available forecasting methods due to their limited flexibility to effectively deal with the extremely wide variety of SKU demand patterns represented in the product portfolio.

This research proposes a novel nonparametric empirical Bayes-based bootstrapping approach (EBBA) as a possible solution for the challenge faced by this distributor and many other companies. We use the firm's actual demand data to compare the forecasting performance of the EBBA method with seven proven methods described in the literature. Experimental results reveal the robustness of the EBBA as it achieves better overall performance than the other methods in terms of inventory efficiency, fewer backorders, and achieving desired customer service levels. EBBA is only constrained from achieving a target customer service level when demand patterns are nonstationary with steep increases in mean and standard deviation of nonzero demand over time. This favorable performance is due to the approach's flexibility to accommodate disparate and complex demand patterns, including interdependent demand intermittency, irregular demand distribution, and even nonstationary demand distribution to some extent.

In summary, the major contributions of the paper are:

1. The paper proposes a novel nonparametric Bayesian bootstrapping method that is highly flexible to accommodate the rich diversity of demand patterns, including extreme complexity, facing spare parts inventory systems, and thus provide robust solutions in terms of inventory efficiency and meeting customer service objectives.
2. The paper introduces a distinctive empirical demand data set for 46,272 truck spare parts that displays extreme demand complexity. The dataset is available upon a proper request to the authors of this study and can be used for checking flexibility and robustness of a forecasting method in dealing with highly complex demand patterns.
3. The paper compares the forecasting performance of the new Bayesian method with that of several proven parametric and nonparametric methods on the empirical dataset, which demonstrates the superior performance of the new method and reveals some new insights into spare parts demand forecasting.

The remainder of this paper is organized as follows Section 2. positions this research in the literature and provides the research background Section 3. presents details of our new Bayesian method for spar part demand forecasting Section 4. investigates the method's performance in comparison with several proven benchmark methods for our empirical dataset Section 5. discusses our findings and concludes the paper with future research suggestions.

2. RESEARCH background

The intermittent demand, spare parts forecasting literature is vast and expanding rapidly. Thus, this survey discusses only research closely related to this project. For a more in-depth literature review, see Boylan and Syntetos (2010), Hu, Chakhar, Siraj and Labib (2018) and Hasni, Aguir, Babai and Jemai (2019), Petropoulos, Makridakis, Assimakopoulos and Nikolopoulos (2014), Syntetos, Babai, Boylan, Kolassa and Nikolopoulos (2016), and most recently Pince, Turrini and Meissner (2021). As common in the literature, we classify the research by parametric and nonparametric methods.

2.1. Parametric methods

Parametric methods assume demand follows a hypothesized probability distribution (e.g., Normal or Poisson) whose parameters are estimated using a forecasting method. The demand distribution parameters are then used to set the inventory policy parameters (e.g., reorder point, order up to level, safety stock) by extrapolating them to predict the RCD distribution. Due to their ease of implementation and ability to provide reasonably good forecasts, the parametric SES and CM are frequently incorporated into Enterprise Resource Planning (ERP) type solutions and forecasting software (Boylan & Syntetos, 2010; Forecast Pro, 2020). SES, initially developed to forecast spare parts demand for faster-moving items, estimates the mean demand without distinguishing among zero and nonzero demand periods. Croston (1972) observes that intermittent demand patterns are constructed from two elements, nonzero demand size and the time interval between demand occurrences, both of which can be independently forecast using exponential smoothing. Croston (1972) assumes nonzero demand sizes are normally distributed, demand occurrences follow a Bernoulli process where inter-arrival times of demand occurrences are geometrically distributed, and nonzero demand sizes and inter-arrival times are mutually independent. The ratio of the predicted nonzero demand size over inter-arrival time provides a point estimate of average demand per unit time, which when combined with an estimate of the variance of forecast errors, can estimate the RCD's CDF of the hypothesized probability distribution. When demand occurs in every time period, the CM and SES forecasts produce identical results.

Syntetos and Boylan (2001) show that the CM is biased with a tendency to over-forecast demand. Syntetos and Boylan (2005) propose the SBA method to nearly eliminate the CM's bias and show, using simulated and empirical data, the SBA more accurately predicts intermittent demand than SES and CM. Teunter, Syntetos and Babai (2011) propose the TSB modification to the CM method. The modification updates the probability of demand occurrence and nonzero demand size at the end of every period, instead of after every nonzero demand occurrence. This better predicts demand after long time intervals with zero demand as commonly associated with spare parts demand, and particularly product obsolescence. Empirical studies reported by Syntetos, Babai and Altay (2012) and Syntetos, Lengu and Babai (2013) suggest that compound Poisson distributions, of which Negative Binomial distribution is one special case, often effectively describe intermittent customer demand patterns for many spare part situations.

Several researchers have developed parametric Bayesian-based forecasting methods (e.g., Hill, 1997; Silver, 1965), but few of them address intermittent demand situations. Aronis, Magou, Dekker and Tagaras (2004) propose the Poisson-Gamma Bayesian (PGB) method assuming demand is Poisson distributed and the prior distribution is the Gamma distribution. The posterior predictive distribution of demand per time period is a Negative Binomial distribution. The method yields a closed-form expression of the lead-time demand distribution, which is used for setting the order up to level, S, of an (S-1, S) inventory system in support a new product line of electronic equipment where historical failure rate data is not available. The PGB method, based on a prior built from aggregate historical data from similar items, yields lower base stock inventory levels than a Bayesian approach using engineering reliability prediction models as the prior distribution. Dolgui and Pashkevich (2008) study a spare parts inventory system with multiple slow-moving items and sparse demand history. Their generalized Bayesian method assumes a population-averaged beta distribution as the prior and the binomial distribution as the likelihood function, leading to a beta-binominal posterior distribution for modeling lead time demand distribution for a group of related items.

Babai, Chen, Syntetos and Lengu (2021) propose an alternative parametric compound Poisson Bayesian (CPB) method assuming demand follows a compound Poisson-Geometric distribution. A comparative study using 7400 theoretically generated demand series and an empirical dataset of approximately 3000 SKUs from the automotive sector reveals that both the PGB and CPB Bayesian methods outperform the SBA method and nonparametric WSS method, proposed by Willemain, Smart and Schwarz (2004), in terms of inventory efficiency and meeting a target customer service level. The CPB method achieves the highest service level, but the PGB approach has fewer backorders.

The general benefit of a Bayesian parametric model is that when the empirical data match precisely the assumed parametric distribution of the model, the model will theoretically report the best model performance over all other competing models (Gönen, Johnson, Lu & Westfall, 2019). Nevertheless, all the parametric methods impose strong probabilistic assumptions on the data and model parameters, which constrain their flexibility, and thus ability to effectively address situations where the demand distributions do not match the specific model assumptions. For instance, PGB explicitly depends on an expert's ability to accurately and reliably determine initial estimates of the model parameters, which are the critical issue for obtaining reasonable demand forecasts (Aronis et al., 2004). However, expert's opinions are unrealistic in many cases, especially for spare parts inventory systems where demand often has an irregular distribution. Even experts with considerable knowledge of past demand histories may not reliably estimate, as *a prior*, the characteristics of the demand distribution, such as its mean demand or an estimated inventory value that is greater than 95% of the cumulative demand during the replenishment cycle. In general, if a parametric model is incorrectly specified, that is, the data significantly violates its underlying model assumptions, it can lead to biased and inconsistent predictions, and ultimately to inappropriate inferences and suboptimal recommendations. (Nambiar, Simchi-Levi & Wang, 2019). Frazier, Robert and Rousseau (2020) and Hong and Martin (2020) show that a misspecified Bayesian model can yield an ill-behaved asymptotic posterior distribution.

2.2. Nonparametric methods

Nonparametric methods often reconstruct the replenishment cycle demand distribution from empirical data using a bootstrapping method. Thus, unlike parametric approaches, nonparametric methods do not rely on a specific distributional assumption so that they are more flexible and model misspecification is less a concern. As such, the nonparametric methods appear to be more suitable for a spare parts portfolio having highly irregular demand patterns as exemplified by our dataset. Efron's (1979) bootstrapping approach randomly resamples, with replacement, values from the item's empirical demand history to generate pseudo-replicate histograms of the replenishment demand distribution. The pseudo-replicates are a valid approximation of the true, unknown distribution that generates the datasets (Efron, 1979, Alfaro, Zoller & Lutzoni, 2003). The pseudo-replicates provide a simple frequency interpretation of the CDF, from which the analyst can determine the inventory control parameters for meeting a specified inventory service level. Furthermore, the bootstrapping approach is explicitly governed by the uniform probability model for resampling, i.e., all data points in the original sample have an equal probability of being selected into a resampling sample. So, it is more flexible and robust than the parametric methods when dealing with perturbations on input data.

The WSS method addresses two main limitations of Efron's bootstrapping approach. First, WSS models autocorrelation between successive demand occurrences as a two-state Markov process. Second, in Efron's method, the data obtained for each bootstrap replication is drawn independently from the sample CDF, which may not contain all possible demand values in the population CDF. WSS rectifies this by 'jittering'; that is, adding random variation to the sample demand values prior to inclusion in the reconstructed empirical distribution. Comparative study using multiple industrial datasets indicates that the WSS method produces more accurate predictions of the cumulative RCD than SES and the CM.

Some improvements to WSS are suggested by Rego and Mesquita (2015) and Zhou and Viswanathan (2011). Viswanathan and Zhou (2008) propose the VZ method that generates nonzero demand occurrences using the historical distribution of the inter-arrival times instead of the two-state Markov model in the WSS method. Zhou and Viswanathan (2011) compare the relative performance of VZ and two variants of the SBA parametric method

(Babai & Syntetos, 2007), reporting mixed results. On randomly generated datasets with long demand history, the VZ heuristic satisfies a specified replenishment cycle service level with less inventory holding and shortage costs than the parametric approaches. However, using industrial demand data, characterized by limited historical demand, the parametric approaches perform better.

While researchers investigate both the frequentist and Bayesian parametric methods for forecasting intermittent demand, to the authors' best knowledge, a Bayesian nonparametric method does not exist for this purpose. From the Bayesian perspective, Efron's bootstrapping approach is analogous to a specialized Bayesian model (Rubin, 1981). Hastie, Tibshirani and Friedman (2017) call this classical bootstrapping methodology, on which the WSS method and its improvements are based, a "poor man's" Bayesian model because it is constrained by the strong uninformative prior assumption that all data points in the observed sample have an equal probability of appearing in a resampling sample. However, one notable problem with this resampling procedure is that it may not well represent the distributional feature of observed data. This is usually of little concern when the data have, or approximately have, a regular standard distribution such as Normal distribution, Negative Binomial distribution, or others. However, if demand is characterized by some irregular distributions, the problem becomes troublesome. For example, one item in our industrial dataset has nonzero demands in 30 out of 81 time periods as listed below in time order, [462, 7, 66, 2, 2, 2, 1, 1, 1, 1, 3, 12, 6, 1, 1, 1, 1, 1, 2,] [10, 6, 2, 2, 6, 4, 1, 1, 4, 2, 3, 3].

Using the classical bootstrapping method we will select a demand size 462 with equal probability of 3.33% (=1/30) as selecting any other record into a resampling sample, despite that this data point appears to be a highly irregular observation relative to the other observations such that it should, more reasonably, have a much lower likelihood than the other demand sizes to occur in the future. As a result, the company may maintain an unreasonably high inventory level for this item based on demand forecasts produced by the WSS approach. For instance, when using the WSS approach to calculate the order-up-to level for this item based on 95% target service level, we estimate the order-up-to level to be 424 with a 100% achieved SL that is significantly above the target service level.

Under the Bayesian framework, it is possible to overcome this limitation of the classical bootstrapping method by replacing the uninformative prior assumption with an informative one that allows more (or less) probable data points to be selected into a resampling sample with higher (or lower) probability. In the standard Bayesian approach, the informative prior assumption must be specified before observing any data. However, in our case, knowledge about the relative likelihood of different data points is acquired only after the data are collected. As a result, the standard Bayesian workflow may not work for our task. Instead, the *Empirical Bayes* (EB) paradigm attributed to Robbins (1955) represents a more feasible approach. In Appendix A, we provide a brief mathematical description of the EB approach that justifies this paradigm both theoretically and technically. Basically, the EB approach differentiates itself from the standard Bayesian approach by estimating the prior distribution of model parameters from observed in-sample data. As such, this approach is a compromise between frequentist and Bayesian approaches. On the one hand, the EB approach models a problem using the standard Bayesian techniques that can produce more effective solutions by incorporating a reasonable prior distribution assumption on model parameters. On the other hand, instead of manually fixing these unknown parameters (i.e., the hyper-parameters) in the prior distribution, EB estimates them separately for different data samples, which enhances the usage of local data. Consequently, the EB model is more flexible than a standard Bayesian model because its prior assumption can be adjusted for different data situations (Casella, 1992). This makes the EB approach particularly appealing for forecasting demand data with irregular distributions. Munyangabo, Waititu and Wanjoya (2019) show that EB methods are efficient data-analysis tools, especially when data have irregular distributions. EB methods have received attention from operational researchers and been used to solve a variety of real-world problems. For example, Quigley, Hardman, Bedford and Walls (2011) develop an EB estimator of the frequency of rare events for risk assessment. Chun (2016) proposes an EB method for predicting the number of conforming items during sequential screening. Quigley, Walls, Demirel, MacCarthy and Parsa (2018) develop an EB approach in support of supply chain managers' decisions on the optimal level of investment for improving quality performance under uncertainty. Eckert, Hyndman and Panagiotelis (2021) use an EB prior in their forecasting of Swiss exports. These application cases consistently show that EB methods can produce model outputs that are more accurate and robust when confronted with high uncertainty and data volatility. We refer interested readers to Carlin and Louis (2000) for a brief overview of EB methods and Casella (1992) for a comprehensive tutorial of EB methods, including theoretical explanations and illustrative examples.

Both parametric and nonparametric methods implicitly assume a time-invariant demand distribution and can perform poorly if the demand distribution is nonstationary. Generally speaking, there are two types of stationarity for time series data, strict stationarity and weak stationarity. If a time series is strongly stationary, its probability distribution should be strictly identical over time. In the case of weak stationarity, a time series may have a time-variant distribution shape but its mean and autocovariance should be finite and unchanged. In this regard, a spare parts portfolio with extremely complex demand patterns is more likely to include nonstationary demand data. Parametric methods subject to strong distribution assumptions are apparently more vulnerable to nonstationary demand distribution. On the other hand, the bootstrapping-based methods may generally be more robust against time-variant demand distribution. This is because their predicted demand distribution for each stocked item is empirically determined by continuously resampling from its own observed demand data, thus more quickly adjusting to a change of demand distribution embodied in newly observed data.

It is worth mentioning that advancements in machine learning and artificial intelligence also motivate the development of neural network (NN) methods for spare parts demand forecasting. Zhang, Patuwo and Hu (1998) provide an overview of NN forecasting methods, including the effect of key factors on forecasting performance. Gutierrez, Solis and Mukhopadhyay (2008), Kourentzes (2013), Lolli et al. (2017), and Babai et al. (2020) are the most prominent research studies using NNs for intermittent demand spare parts forecasting. Unlike traditional methods, NNs are nonlinear data-driven, self-adaptive methods with the ability to learn from data samples and identify hidden functional relationships among data without any distribution assumptions. However, their effectiveness for spare parts forecasting and inventory control in industrial settings is unproven at this time. Gutierrez et al. (2008) note that their NN-GUT method consistently outperforms SBA, CM, and SES methods when the average nonzero demand size is smaller for the training sample as compared to the test samples, but usually underperforms when the average demand size is greater for the training samples. In other words, the performance of NN forecasting methods is also constrained by nonstationary demand distribution. Babai et al. (2020) propose several modifications to the NN-GUT method and compare the performance of three variants of their NN-LAG model against the SES, SBA, CM, WSS, VZ, NN-GUT, assuming a periodic (T,R) order-up-to inventory system with a cycle service level (CSL) objective. The results are inconclusive in determining which method attains the desired CSL at the lowest inventory investment. Boylan and Syntetos (2021) note the difficulty of developing NNs for intermittent demand forecasting and the extra caution needed to study their predictive performance. Hence, in accordance with the main scope of this paper, we do not include a NN method in this study.

2.3. Comparative studies

Overall, research seeking to demonstrate the superiority of any particular parametric or nonparametric approaches to forecasting spare parts with intermittent demand is inconclusive. Syntetos et al. (2015) compare the performance of WSS, SES, CM, and SBA using inventory efficiency curves finding marginal superiority of WSS over the parametric methods when lead times are short, and demand is moderately irregular. However, SBA is the best performer given more irregular demand and longer lead times. Considering the SBA, VZ, and WSS methods, Hasni et al. (2019) show that forecasting method performance is sensitive to the problem's length of demand history, replenishment cycle lead time, demand distribution parameters, and cost parameters. Research findings, using computer generated demand, reveal: (i) SBA has lower inventory cost than VZ for problems with higher coefficient of variations for demand interval and/or nonzero demand, (ii) for long demand histories, VZ is the best performing method if the variability of nonzero demands is low, but SBA outperforms the bootstrapping approaches as variability increases. For highly variable nonzero demands, WSS outperforms VZ, but not SBA, and (iii) for short demand histories, SBA is dominant for short lead times, VZ is best with low nonzero demand variability, and WSS is preferred only when nonzero demand variability and lead times are high. Using empirical datasets on 5000 aircraft spare parts, SBA leads to the lowest inventory holding and backorder cost in 82.4% of the test problems. As expected, the estimator with the highest inventory level coincides with the highest fill rate, and visa-versus. In a comprehensive survey of intermittent demand forecasting methods, Pince et al. (2021) find that SBA and CM seem to be better choices for industry application than other parametric methods, and the WSS method is usually superior to the other nonparametric methods in terms of accuracy (57% to 43%) and inventory (63% to 28%) measures. In general, the nonparametric methods outperform parametric methods on accuracy measures (64% to 35%) and inventory measures (60% to 33%), but they conclude that it is difficult to provide clear-cut recommendations between parametric and nonparametric methods in terms of inventory performance and empirical data type. This difficulty is at least partially due to limited diversity of the demand patterns in the empirical and simulated datasets used by the existing studies. Hence, specific models perform best on datasets aligned with their model assumptions, but they have limited flexibility to address the variety of complex demand patterns beyond the model assumptions.

We contribute to the body of research by developing and testing an empirical Bayes-based bootstrapping approach (EBBA) that is structurally analogous to the WSS bootstrap heuristic. However, it is considerably more flexible in dealing with extremely irregular demand patterns and strongly interdependent demand intermittency. Using a large empirical dataset, we evaluate and compare our approach's performance against the most proven nonparametric and parametric approaches and a comparable Bayesian method (CPB) in the literature. Our analysis shows that the new Bayesian method is superior to the other methods in terms of overall forecasting accuracy and inventory performance for the large spare parts portfolio that displays extremely high complexity in demand patterns.

3. MODEL description

We propose a unique nonparametric Empirical Bayesian bootstrap approach (EBBA) to forecast spare parts demand. The method accommodates interdependent demand intermittency and irregular nonzero demand to overcome the weaknesses of the existing approaches and provide a better foundation for inventory control.

Following Willemain et al. (2004), an intermittent demand dataset $\boldsymbol{D}_T = (D_1, D_2, \ldots, D_T)$ is decomposed into two separate components. One is a two-state Markov chain, denoted by $\boldsymbol{M}_T = (M_1, M_2, \ldots, M_T)$, of sequential 0–1 state indicator values. The other has assumingly nonzero demand sizes in all periods, denoted by $\boldsymbol{D}_T^+ = (D_1^+, D_2^+ \ldots, D_T^+)$. Both $\boldsymbol{M}_T$ and $\boldsymbol{D}_T^+$ are simulated from posterior distributions derived from the proposed Bayesian models. The forecast of the actual demand series is the Hadamard product of $\boldsymbol{M}_T$ and $\boldsymbol{D}_T^+$, $\boldsymbol{D}_T = \boldsymbol{M}_T \circ \boldsymbol{D}_T^+$; that is, all nonzero demand sizes in $\boldsymbol{D}_T$ are randomly intermitted by the "0″ state in the accompanying $\boldsymbol{M}_T$.

3.1. Bayesian inference of the time interval between demand occurrences

Let $\boldsymbol{M}_T$ be a time homogenous binary Markov chain of 0–1 state indicator values with the one-step probability transition matrix

$$\boldsymbol{P}_{tr} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix},$$

where $p_{00} = 1 - p_{01}$ and $p_{11} = 1 - p_{10}$. It is worth noting that although $\boldsymbol{P}_{tr}$ is time independent, the n-step transition matrix $\boldsymbol{P}_{tr}^{(n)} = \boldsymbol{P}_{tr}^n$ is generally dependent on the time steps $n \geq 2$. The likelihood function of p_{01}, p_{10} is

$$L(\boldsymbol{M}_T |\ p_{01}, p_{10}) = {p_{00}}^{t_{00}} p_{01}^{t_{01}} p_{10}^{t_{10}} {p_{11}}^{t_{11}} = (1 - p_{01})^{t_{00}} p_{01}^{t_{01}} p_{10}^{t_{10}} (1 - p_{10})^{t_{11}}$$

where t_{ij} is the number of one-step transition from state i to j in $\boldsymbol{M}_T$, and $\sum_{i,j=0,1} t_{ij} = T - 1$. The inference on p_{01}, p_{10} can follow either a frequentist or Bayesian approach. Within the frequentist framework, the maximum likelihood estimate of $\boldsymbol{P}_{tr}$ is

$$\hat{\boldsymbol{P}}_{tr} = \begin{bmatrix} \hat{p}_{00} & \hat{p}_{01} \\ \hat{p}_{10} & \hat{p}_{11} \end{bmatrix} = \begin{bmatrix} \frac{t_{00}}{t_{00}+t_{01}} & \frac{t_{01}}{t_{00}+t_{01}} \\ \frac{t_{10}}{t_{10}+t_{11}} & \frac{t_{11}}{t_{10}+t_{11}} \end{bmatrix}.$$

The k-step forecast of demand states, $M_{T+1}, \ldots, M_{T+k}$, are simulated from $\hat{\boldsymbol{P}}_{tr}$ conditioning on the current state M_T. One drawback of this procedure is that it only provides pointwise estimates of p_{01}, p_{10} and thus, does not account for the full uncertainty about the two parameters. For instance, if one historical (i.e., sampled) dataset does not include any periods with zero demand, the frequentist estimate unreasonably rules out the possibility of future periods with zero demand. In other words, $\hat{p}_{01} = \hat{p}_{10} = 0$.

$$\hat{\boldsymbol{P}}_{tr} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

In this case, a properly chosen parametric model (e.g., compound Poisson distribution) may perform better than the WSS approach if the data appear to be generated from the assumed distribution.

The Bayesian approach infers the posterior distribution of p_{01}, p_{10}, $f(p_{01}, p_{10}|\boldsymbol{M}_T)$, by updating the prior distribution $\pi(p_{01}, p_{10})$ using observed values $\boldsymbol{M}_T$. Hence, a potential benefit of the Bayesian estimation over its frequentist counterpart is that it considers the full demand distributional form, and thus full uncertainty about the values of p_{01}, p_{10}. Following the above example, even if a historical dataset includes all nonzero demands, the Bayesian posterior samples of p_{10} still contains positive $p_{10} > 0$. That implies there is still a possibility of observing zero demand in the future, representing a more reasonable parameter estimation.

As a result, this Bayesian treatment can better deal with a change in demand intermittency pattern over time.

To obtain the distributional estimates, we must specify the prior distribution $\pi(p_{01}, p_{10})$, but there are not any set rules governing how a prior distribution *should* be placed on the model parameters. As such, prior specification is essentially subjective. In Bayesian inference for a probability parameter p, the conventional choice is to assign it a Beta distribution, $\pi(p) \sim \text{Beta}(\alpha,\ \beta)$. For the bivariate parameter vector (p_{01}, p_{10}), a natural extension of the Beta prior is the bivariate Dirichlet prior $\pi(p_{01}, p_{10}) \sim \text{Dir}(\alpha_{01}, \alpha_{10})$. But, because it requires $p_{01} + p_{10} = 1$, the Dirichlet prior implicitly assumes the n-step transition matrix $\boldsymbol{P}_{tr}^{(n)}$ is time-independent (i.e., $\boldsymbol{P}_{tr}^{(n)} = \boldsymbol{P}_{tr}^{n} = \boldsymbol{P}_{tr}$). This assumption is rather unrealistic for most real-world scenarios. Another possible alternative is to assign a Beta prior (e.g., $\text{Beta}(0.5,\ 0.5)$) separately to p_{01} and p_{10}. However, this prior imposes a strong presumption that p_{01} and p_{10} are independent of each other, which may not reflect the actual situation in our problem.

Instead, we use a Jeffrey's prior on p_{01}, p_{10}, which is defined in terms of the Fisher information

$$\pi(p_{01}, p_{10}) \propto\ I(p_{01}, p_{10})^{1/2}$$

where the Fisher information $I(p_{01}, p_{10})$ is given by

$$I(p_{01}, p_{10})^{1/2}\ = E\left[-\frac{\partial^2 L(\boldsymbol{M}_T|\ p_{01}, p_{10})}{\partial p_{01} \partial p_{10}}\right].$$

A simple functional form of the Jeffrey's prior, derived in Assoudou and Essebbar (2004), is

$$\pi(p_{01}, p_{10}) \propto\ \left[(T-1)(p_{01}+p_{10}) - 1 + (1-p_{01}-p_{10})^{T-1}\right]^{\frac{1}{2}}$$

$$\times \left[(T-1)p_{01}(p_{01}+p_{10}) + p_{10}(1-p_{01}-p_{10})^{T-1}\right]^{\frac{1}{2}}$$

$$\times\ p_{01}^{t_{01}-1/2}(1-p_{01})^{t_{00}-1/2} p_{10}^{t_{10}}(1-p_{10})^{t_{11}-1/2}(p_{01}+p_{10})^{-2}. \quad (1)$$

With this prior setting, the model doesn't rely on either of the two aforementioned assumptions, thus suggesting a better fit for real-world data. The Bayesian posterior distribution $f(p_{01}, p_{10}|\boldsymbol{M}_T)$ for the Jeffrey's prior can be easily estimated, for example, by applying the Independent Metropolis-Hastings (IMH) algorithm (Robert & Casella, 2013).

3.2. Bayesian bootstrapping inference of nonzero demand size

The WSS approach forecasts the nonzero demand size by randomly sampling, with replacement, from the historical set of nonzero demands. The most obvious advantage of this method is that it does not depend on standard distributional assumptions. Instead, all observations in the historical dataset are equally likely to be selected according to a uniform resampling distribution, with replacement. While this assumption is intuitively justified under the frequentist framework, it is less reasonable in the Bayesian domain.

The Bayesian analogy to the classical bootstrapping methodology assumes a prior distribution of resampling probability as $\boldsymbol{\pi_{prior}} \sim \text{Dirichlet}(\boldsymbol{\alpha_n})$, which is a Dirichlet distribution with the hyperparameter vector $\boldsymbol{\alpha_n} = [\alpha_1, \alpha_2 \ldots, \alpha_n\ ]$ where n is the size of the observed data sample (in our case, the sample of all nonzero demand data). The posterior distribution can be given as $\boldsymbol{\pi_{posterior}} \sim \text{Dirichlet}(1_{\boldsymbol{n}} + \boldsymbol{\alpha_n})$, where $1_{\boldsymbol{n}}$ is an all-ones vector $[1_1, 1_2, \ldots, 1_n]$. To see it, simply consider that an observed nonzero demand sample of size n, $\boldsymbol{d}^+ = (d_1^+, d_1^+ \ldots,\ d_n^+)$, is one realization of random resampling on its own. The parameter vector $\boldsymbol{\pi} = (\pi_1,\ \ \pi_2, \ldots, \pi_n)$ is the probabilities of each demand data point being selected into this resampling sample, whose likelihood function conditional on the sample is

$$L\left(\boldsymbol{\pi}|\boldsymbol{d}^+\right) \propto \prod_{i=1}^{n} \pi_i,$$

which is the multinomial likelihood function. We assign $\boldsymbol{\pi_{prior}} \sim \text{Dirichlet}(\boldsymbol{\alpha_n})$ because it is conjugate to the multinomial data distribution. Then, the posterior distribution of $\boldsymbol{\pi}$ is given as

$$\boldsymbol{\pi_{posterior}} \propto L\left(\boldsymbol{\pi}|\boldsymbol{d}^+\right)\boldsymbol{\pi_{prior}} \propto \prod_{i=1}^{n} \pi_i^{1+\alpha_i-1},$$

which is $\text{Dirichlet}(1_{\boldsymbol{n}} + \boldsymbol{\alpha_n})$. This classical bootstrapping method is a special case of the general Bayesian model in which we let $\boldsymbol{\alpha_n} \to 0_{\boldsymbol{n}} = [0_1,\ 0_2 \ldots,\ 0_n\ ]$ (Hastie et al., 2017, p.272) and thus use a strong uninformative prior distribution $\boldsymbol{\pi_{prior}} \sim \text{Dirichlet}(0_{\boldsymbol{n}})$ to do the resampling. In this limit when $\boldsymbol{\alpha_n} = 0_{\boldsymbol{n}}$, we concentrate the probability mass on one data point in the sample at random indiscriminately. Thus, all sample data points have an equal probability of appearing in the new sample.

We, instead, estimate $\boldsymbol{\alpha_n}$ based on the empirical probability density function (EPDF) values of observed nonzero demand sizes. Denoting $\hat{\boldsymbol{f}}(\boldsymbol{d}_n^+) = [\hat{f}(d_1^+),\ \hat{f}(d_2^+),\ \ldots,\ \hat{f}(d_n^+)]$ as the vector of EPDF values computed at the observed nonzero demand sample $\boldsymbol{d}_n^+$, we assume $\boldsymbol{\alpha_n}$ has the value in the form:

$$\boldsymbol{\alpha_n} = \hat{\boldsymbol{f}}\left(\boldsymbol{d}_n^+\right)/\hat{\boldsymbol{f}}\left(\boldsymbol{d}_n^+\right)_1$$

so that

$$\boldsymbol{\pi_{prior}} \sim \text{Dirichlet}\left(\hat{\boldsymbol{f}}\left(\boldsymbol{d}_n^+\right)/\hat{\boldsymbol{f}}\left(\boldsymbol{d}_n^+\right)_1\right) \quad (2)$$

where $\hat{\boldsymbol{f}}(\boldsymbol{d}_n^+)/\hat{\boldsymbol{f}}(\boldsymbol{d}_n^+)_1$is $\hat{\boldsymbol{f}}(\boldsymbol{d}_n^+)$ scaled by its L^1- norm. For the sample item discussed in Section 2.2, this prior distribution suggests that we should select the demand size of 462 into a resampling sample with only a 0.26% probability, significantly lower than the uniform probability of 3.33% (= 1/30) as in the classical bootstrapping. On the other hand, the classical bootstrapping will select demand size of 1 with 36.7% (= 11/30) probability, which is lower than the probability of 48.18% (= 4.38% × 11) as suggested by the $\boldsymbol{\pi_{prior}}$ in (2). Alternatively, the Bayesian bootstrapping assuming this prior distribution yields the order-up-to level for the same term based on a 95% targeted CSL to be 8 with a 95.83% achieved CSL that matches the target rate. So the prior distribution provides a more reasonable resampling procedure for spare part demand data because demand sizes with higher (or lower) empirical density values are more likely (or less likely) to appear in resampling.

In essence, EBBA can be considered as an improvement to WSS that significantly enhances its flexibility to deal with interdependence of demand intermittency and irregular nonzero demand distribution. The implementation procedure for the proposed Bayesian method is summarized below.

Bayesian Forecasting Procedure

1: Separate an observed demand dataset $\boldsymbol{d}_T$ into the two-state Markov chain part $\boldsymbol{m}_T$ and the all-positive demand part $\boldsymbol{d}_n^+$, $n \le T$.
2: Repeat Steps 3–6 N times to produce posterior samples of k-step demand forecast values
3: Simulate one posterior transition matrix conditional on $\boldsymbol{m}_T$ and the Jeffrey's prior in (1).
4: Generate one k-step two-state Markov chain forecast $\boldsymbol{m}_k^f$ given the current state using the posterior matrix simulated in Step 3.
5: Repeatedly select k values from $\boldsymbol{d}_n^+$ based on the prior resampling distribution in (2) and jitter the selected values based on Equation (7) in Willemain et al. (2004) to be the nonzero demand forecast values $\boldsymbol{d}_k^{f+}$.
6: Compute the Hadamard product of $\boldsymbol{m}_k^f$ and $\boldsymbol{d}_k^{f+}$ as one posterior sample of k-step demand forecast values.
7: Sum the k forecast values in each of the N posterior samples to provide posterior samples of replenishment cycle demand; i.e., k = Lk=L for the continuous review, fixed order quantity (Q, r) model and $k = T + \text{L}$ for the periodic review (T, R) inventory model.

Table 1
Summary Bi-weekly Statistics of Truck Spare Parts Industry.

	Demand interval		Demand size		Demand per period	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Minimum	1.00	0.00	1.05	0.21	0.22	0.50
25th Percentile	1.01	0.12	3.08	2.30	1.64	2.24
Median	1.26	0.59	6.54	5.22	4.29	4.97
75th Percentile	1.84	1.30	19.78	14.59	15.49	14.13
Maximum	6.09	12.91	69,481.98	253,421.14	62,433.09	253,421.14

4. COMPARATIVE assessment of model performance

The comparative assessment of forecasting and inventory performance includes EBBA and the most proven methods in the literature; parametric methods include SES, CM, SBA, TSB, and CPB, and nonparametric methods include WSS and VZ. A description of the forecasting methods and the procedural steps for each method are given in Appendix B. Since WSS, VZ and EBBA estimate the entire RCD distribution, traditional point estimate error metrics, such as the mean absolute deviation (MAD) of forecast error, fall short in assessing the accuracy of the nonparametric forecasts. We overcome this difficulty by evaluating forecast accuracy and performance on two dimensions: (i) pooled percentile estimates of the RCD distribution and (ii) inventory performance in a periodic review order-up-to inventory system.

In practice, EBBA provides considerable statistical information that is not available from parametric estimation methods. The approach is highly flexible and robust to accommodate varying intermittent and irregular demand features without prior knowledge of demand characteristics. It is equally applicable in continuous and periodic review inventory systems. EBBA implementation is straightforward with reasonable CPU computing times. This study was run using parallel computing and performed in R by running an instance with 48 v CPUs on the Amazon Elastic Compute Cloud (EC2). It took less than 4 hours to process all the 46,272 demand datasets. Considering the rapid advancements in business analytics computing, these times would be considerably lower on high performance systems and are expected to fall rapidly over time. We illustrate one sample forecast output of EBBA in Appendix C.

4.1. Industrial dataset

We assess forecasting performance using the empirical truck spare parts data described earlier in the paper. The test datasets, selected in consultation with company executives and subject matter experts, include approximately three years of bi-weekly demand data for 46,272 SKUs that are stocked at the firm's largest distribution center. The SKUs are a subset of 85,900 SKUs satisfying the following criteria: (i) demand exceeds a total of 30 units during the most recent three years, (ii) at least 20 percent of the 81 bi-weekly time periods have nonzero demand occurrences and (3) there are at least two nonzero demand points in the first 56 bi-weekly periods.

Table 1 reports some summary statistics of the data. In Appendix D, Fig. D1 illustrates the actual demand and kernel estimates of the empirical probability distributions for six typical SKUs. The table and figure reveal two novel features of our demand dataset. First, the time periods with zero or non-zero demand often appear consecutively in clusters, suggesting strong interdependence of demand intermittency. Second, the nonzero demand has extremely irregular distributions (e.g., multimodal, flat-topped, extremely long-tailed, relatively symmetric, or nonstationary). The extreme irregularity of the demand distributions is most often a consequence of frequent but unexpected occurrence of certain demand sizes, which cannot be considered as the "normal" abnormal observations under a probability distribution of standard form.

4.2. Forecast distributional accuracy assessment

As in Willemain et al. (2004), we note that the nonzero demand occurrences are sparse for each SKU, which suggests using the last replenishment cycle's k-period demand forecast for evaluation. We use the item's earlier demand periods as the in-sample periods for estimating the k-step RCD posterior samples for predicting the RCD distribution based on some forecasting method. For ith SKU in our dataset, $i = 1, 2, \ldots, 46,272$, denote $F_{RCD}(D)_i$ as the cumulative RCD distribution predicted by one forecasting method. We compute the probability of no stock out for this SKU during the last order cycle as $F_{RCD}(d_i)_i$, where d_i is the holdout sample's demand value. As a result, we obtain a large sample set of probabilities of no stock out $P_D = \{F_{RCD}(d_1)_1, F_{RCD}(d_2)_2, \ldots, F_{RCD}(d_{46,272})_{46,272}\}$. According to the universality of the uniform, if the forecasting method predicts all $F_{RCD}(D)_i$s correctly, P_D should follow a standard uniform distribution. That is that the 20% percentile of P_D is delineated at 0.2, the median at 0.5, and 99% percentile at 0.99. A higher degree of conformance of the estimated percentiles to the standard uniform distribution is an indicator of a more accurate estimation of the RCD distribution by one forecasting method.

Willemain et al. (2004) find the distributional accuracy of the WSS method is higher than the SES and CM heuristics for replenishment cycle lengths of 1, 3 and 6 time periods. We expand the comparison to include eight forecast methods. For CM, SBA, TSB, and SES, we study situations where demand is assumed to be normally distributed as suggested by Willemain et al. (2004), or negative binomially distributed as in Babai et al. (2021). In Appendix E, Table E1 presents the estimated percentiles of the probability set P_D inferred from the different forecasting methods. The results show that, overall, the EBBA percentile estimates have a higher degree of conformance to the uniform distribution than the others. But EBBA consistently overestimates the high percentiles. Accordingly, this means EBBA tends to underestimate the probability of larger demand size, which is expected for our proposed method. When one SKU has high variation in its demand size, especially because of some change in demand distribution over time, its empirical demand distribution (which governs the resampling procedure in EBBA) estimated from one observed demand data sample can be inaccurate. Meanwhile, every SKU's nonzero demand size is naturally bounded by zero from below and unbounded from above. In consequence, the inaccurate inference of empirical demand distribution is most likely reflected in its underestimation of the probability of larger demand size. We observe that the assumption of Negative Binomial demand distribution generally works better than the Normal distribution assumption, which is consistent with the findings in the literature (Syntetos et al., 2013, 2012). The relatively inferior performance of CPB is not a surprise. This result demonstrates how model misspecification can affect one parametric model's forecasting performance. CPB assumes customer demand over n periods is a Levy process with independent and iden-

Table 2
Summary Experimental Results for all SKUs.

	Target Service Level 99%							
	Stock on-hand		Backorder		FR (%)		CSL (%)	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
CM	18.62	188.06	0.00	16.48	100.00%	96.08%	100.00%	96.58%
SBA	18.5	205.72	0.04	9.72	98.95%	95.69%	96.00%	96.01%
TSB	18.06	189.44	0.05	55.60	98.94%	95.33%	96.00%	95.46%
SES	29.8	252.70	0.00	10.13	100.00%	96.34%	100.00%	96.82%
WSS	25.88	259.01	0.00	2.73	100.00%	97.99%	100.00%	98.23%
VZ	20.94	246.99	0.00	6.34	100.00%	96.27%	100.00%	95.14%
CPB	28.02	71.63	0.00	82.14	100.00%	93.77%	100.00%	88.13%
EBBA	15.44	90.25	0.00	7.84	100.00%	95.31%	100.00%	94.85%
	Target Service Level 95%							
	Stock on-hand		Backorder		FR (%)		CSL (%)	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
CM	11.55	107.40	0.24	18.28	95.60%	92.29%	92.00%	92.61%
SBA	10.70	109.16	0.36	12.27	93.93%	91.03%	92.00%	90.79%
TSB	10.56	105.74	0.36	57.70	94.00%	90.65%	92.00%	90.13%
SES	15.20	123.21	0.24	12.37	95.65%	92.31%	92.00%	92.94%
WSS	16.34	105.16	0.00	6.57	100.00%	95.89%	100%	95.89%
VZ	13.44	94.77	0.04	12.91	99.20%	93.00%	96.00%	90.60%
CPB	10.90	46.85	0.12	92.10	95.65%	87.59%	96.00%	77.20%
EBBA	9.42	63.27	0.24	13.26	95.92%	90.91%	92.00%	89.23%
	Target Service Level 90%							
	Stock on-hand		Backorder		FR (%)		CSL (%)	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
CM	8.64	81.86	0.48	20.49	89.49%	86.44%	88.00%	88.24%
SBA	7.64	80.37	0.68	16.09	91.86%	88.70%	88.00%	85.16%
TSB	7.60	79.40	0.68	60.15	89.35%	86.67%	88.00%	84.67%
SES	10.38	88.97	0.52	15.30	91.67%	88.22%	92.00%	88.79%
WSS	11.96	80.74	0.08	9.66	98.25%	93.41%	96.00%	92.36%
VZ	9.63	71.09	0.24	18.11	95.00%	89.63%	92.00%	86.45%
CPB	7.72	42.66	0.44	94.12	88.24%	84.12%	92.00%	73.04%
EBBA	7.16	52.67	0.48	16.37	92.32%	87.64%	88.00%	84.85%
	Target Service Level 85%							
	Stock on-hand		Backorder		FR (%)		CSL (%)	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
CM	7.00	68.52	0.68	22.67	85.71%	82.91%	88.00%	84.16%
SBA	5.98	66.17	0.96	19.69	88.71%	85.61%	80.00%	80.00%
TSB	6.00	65.74	0.96	62.55	85.61%	82.99%	80.00%	79.71%
SES	8.18	73.68	0.76	18.08	88.35%	84.97%	84.00%	85.16%
WSS	9.50	67.39	0.24	12.19	95.45%	90.99%	92.00%	89.31%
VZ	7.56	59.66	0.48	20.54	91.43%	86.64%	88.00%	82.92%
CPB	6.26	40.73	0.72	94.99	83.68%	81.38%	84.00%	70.00%
EBBA	5.90	45.78	0.72	18.73	89.30%	85.00%	84.00%	81.28%

tically distributed (i.i.d) increments. However, demand intermittency in our data shows a strong correlation between different periods. Ignoring the strong data interdependence in demand forecasting is especially concerning in that it may accumulate forecasting error over time. The assumption that demand follows an identical parametric distribution in every period is also not valid for our dataset. As a result, the posterior predictive demand distribution for many SKUs in our dataset can be ill-behaved.

4.3. Assessing forecast and inventory performance of the forecasting methods

We conduct a comparative assessment of the inventory efficiency and achieved customer service level performance among the forecast methods. The experiments use the datasets for the 46,272 SKUs of the automotive parts distributor. The first 56 time periods are the in-sample data for establishing the forecast method parameter values and the prior distribution.[1] The remaining 25 periods constitute the out-sample data for evaluating the performance of the forecasting methods.

As suggested by Gardner and Koehler (2005), Syntetos et al. (2015), Turrini and Meissner (2019), and Babai et al. (2021), we measure forecast performance based on its impact on the firm's inventory system efficiency in terms of average inventory, backorder levels, fill rate and achieved customer service level. We simulate the periodic review (TR) inventory system with order-up-to level R. The review interval T and lead time L are both set equal to one bi-weekly time period, equating to a 4-week replenishment cycle, as is common for the majority of the firm's SKUs. R is updated for each out-of-sample period to meet a target replenishment cycle service level (CSL), as defined by the probability of no stock-outs during a replenishment cycle. Forecast method performance is evaluated at CSL targets of 85%, 90%, 95% and 99%. For each of these service levels, we also calculate the associated fill rate (FR) as defined by the fraction of demand that is served by stock on-hand, which is a widely accepted customer service metric in industry (Guijarro, Cardos & Babiloni, 2012; Teunter, Syntetos & Babai, 2017).

Due to its support by empirical evidence (i.e., Babai et al., 2021; Syntetos & Boylan, 2006; Syntetos et al., 2015), the parametric forecast methods (CM, SBA, TSB, and SES) model lead-time demand using the Negative Binomial distribution (NBD) to esti-

[1] For the parametric methods, the demand of the first 12 periods are used as an initialization sample to computer the beginning nonzero demand size and demand interval. If no demand occurs in the first 12 time periods, the initial demand interval is set as 1 for CM, SBA, and TSB.

Table 3
Summary Experimental Results for the Subset with Superior EBBA Performance (Top 80%).

Target Service Level 99%					Target Service Level 95%				
	Stock on-hand	Backorder	FR (%)	CSL (%)		Stock on-hand	Backorder	FR (%)	CSL (%)
CM	112.27	3.14	97.13%	97.69%	CM	72.2	3.56	93.61%	94.26%
SBA	119.51	0.89	96.73%	97.09%	SBA	72.77	1.59	92.34%	92.51%
TSB	111.46	19.33	96.53%	96.77%	TSB	71.5	19.93	92.08%	92.03%
SES	135.32	0.92	97.27%	97.84%	SES	82.73	1.38	93.77%	94.67%
WSS	138.03	0.19	99.02%	99.40%	WSS	83.55	0.32	97.86%	98.29%
VZ	128.23	1.85	97.65%	97.11%	VZ	75.09	3.14	95.16%	94.13%
CPB	61.66	33.03	94.49%	92.97%	CPB	37.06	39.21	91.42%	84.08%
EBBA	73.49	0.35	97.80%	98.33%	EBBA	51.59	1.07	94.31%	94.26%
Target Service Level 90%					Target Service Level 85%				
	Stock on-hand	Backorder	FR (%)	CSL (%)		Stock on-hand	Backorder	FR (%)	CSL (%)
CM	57.44	4.11	90.05%	90.28%	CM	49.43	4.78	86.89%	86.46%
SBA	56.89	2.74	87.98%	87.38%	SBA	48.57	3.88	84.21%	82.56%
TSB	56.7	20.58	87.87%	87.01%	TSB	48.64	21.32	84.24%	82.37%
SES	64.49	2.02	89.97%	91.09%	SES	55.17	2.71	86.76%	87.78%
WSS	65.48	0.54	96.05%	96.30%	WSS	55.09	0.87	93.97%	93.94%
VZ	57.45	4.78	92.28%	90.99%	VZ	48.15	5.22	89.46%	87.95%
CPB	32.68	40.57	88.06%	80.32%	CPB	30.64	41.17	85.20%	77.32%
EBBA	42.55	1.81	91.19%	90.57%	EBBA	36.03	2.48	88.52%	87.33%

mate the order-up-to level. The order-up-to level for period t is $R_t = \phi^{-1}_{L+T,t}(FR)$, where $\phi^{-1}_{L+T,t}(\cdot)$ is the inverse of the cumulative NBD of demand during the replenishment cycle (Babai et al., 2021). The mean and variance of NBD are calculated as $(L+T)\cdot F_t$ and $(L+T)\cdot MSE_t$, where F_t and MSE_t are the forecast and smoothed mean squared forecast error calculated for period t. The NBD requires the variance to be greater than the mean. Thus, if the variance is less than the mean, we set the variance at 1.05 times the mean (Babai et al., 2021). For the CPB method, we follow Babai et al. (2021) to model lead-time demand using the Poisson-Geometric distribution and obtain the RCD distribution based on the procedures illustrated in Appendix B with 1000 sampling replications. For the nonparametric methods (EBBA, WSS, and VZ), the RCD distribution is generated using the empirical data with 1000 sampling replications.

The sequence of events in each period is as follows: inventory position is observed, a replenishment order is submitted, the incoming replenishment order placed is received, and customer demand is filled (Cachon and Terwiesch, 2017). The forecast and order-up-to level are updated at the end of the period for use in period $t+1$. The opening and closing inventory are calculated as follows: $open_t = close_{t-1}$ and $close_t = open_t + delivery_t - demand_t$. Correspondingly, on-hand opening inventory is $open^+_t = \max(open_t, 0)$, closing on-hand inventory is $close^+_t = \max(close_t, 0)$. Inventory shortage in period t, is $s_t = max(demand_t - (open^+_t + delivery_t), 0)$. The average stock on-hand during period t is $i_t = (open^+_t + close^+_t)/2$, and the backorder $b_t = \max(-close_t, 0)$. $CSL_t = 1$ if a stock-out does not occur in period t, otherwise $CSL_t = 0$. The average fill rate $\overline{FR}$, achieved $\overline{CSL}$, inventory or stock on-hand $\overline{SOH}$, and number of backorders $\overline{BO}$ are determined for the N out-of-sample periods. Specifically, $\overline{FR} = 1 - \frac{\sum_{t=1}^{N} s_t}{\sum_{t=1}^{N} demand_t}$, $\overline{CSL} = \frac{1}{N}\sum_{t=1}^{N} CSL_t$, $\overline{SOH} = \frac{1}{N}\sum_{t=1}^{N} i_t$, and $\overline{BO} = \frac{1}{N}\sum_{t=1}^{N} b_t$. The objective of the forecast and inventory simulation is to minimize the average stock on-hand subject to meeting the target CSL.

Table 2 summarizes the experimental results of the full product portfolio. There is a considerable discrepancy between the median and mean values of the performance measures, suggesting that the SKUs have skewed demand distributions. Overall, the experimental results support some findings reported in the literature. As reported in Rego and Mesquita (2015), WSS maintains, on average, more stock on-hand resulting in few backorders and higher fill rates and achieved CSL than the parametric methods. Consistent with Hasni et al. (2019), the VZ method has higher average stock on-hand and fill rates than the SBA method. The results also support Syntetos et al. (2015) where the CM achieves higher average CSLs than the SBA method.

A disconcerting finding is that all the experimental forecasting methods had difficulty in achieving the target CSLs. WSS was by far the best performer on this dimension. This finding is not surprising. As illustrated in Section 2 for WSS, the occurrence of a single irregularly large demand size can dramatically lift its required stock on-hand, enabling it to achieve a higher CSL in the presence of highly complex demand patterns. The SES and CM methods were the second and third-best performers, respectively. The parametric CPB is the standout worst performer on achieved CSL. TSB and EBBA methods round out the bottom three CSL performers. As expected, the WSS, the best CSL performer, ranked in the bottom half of the methods in terms of average stock on-hand. Similarly, CPB, the worst CSL performer, maintained the lowest average stock on-hand, but it had by far the highest number of backorders across all methods. EBBA had the second-lowest average stock on-hand but ranked in the top half of the methods with fewer backorder numbers. The experimental methods' backorder and fill rate performances are as anticipated based on their mean respective inventory levels and achieved CSLs.

Overall, the EBBA method's inventory efficiency is encouraging, but its effectiveness in achieving target CSLs compromises its application in practice. This raises the question: Does EBBA systemically underperform on meeting a target service level or is its underperformance attributed to specific demand characteristics associated with a subset of the SKUs in the dataset? If so, is there an identifiable demand pattern that compromises EBBA's performance? We investigate these questions in the following section.

4.4. Robustness of EBBA's performance

As seen in Table 2, EBBA's achieved CSL performance ranges from approximately four to six percent below target with significantly lower stock on-hand. We seek to better understand EBBA's performance by examining its performance at the 99% CSL in more detail. We begin by ranking the experimental results at 99% target CSL for all the SKUs according to EBBA's performance from best to worst in the ordered sequence on three metrics: achieved CSL, stock on-hand, and backorder. We do a multiple-level sorting using achieved CSL in descending order as the first level, stock on-hand in ascending order as the second level, and backorders in ascending order as the third level. Based on the sorted simulation results, we organize the top-ranked 37,000 SKUs (approximately

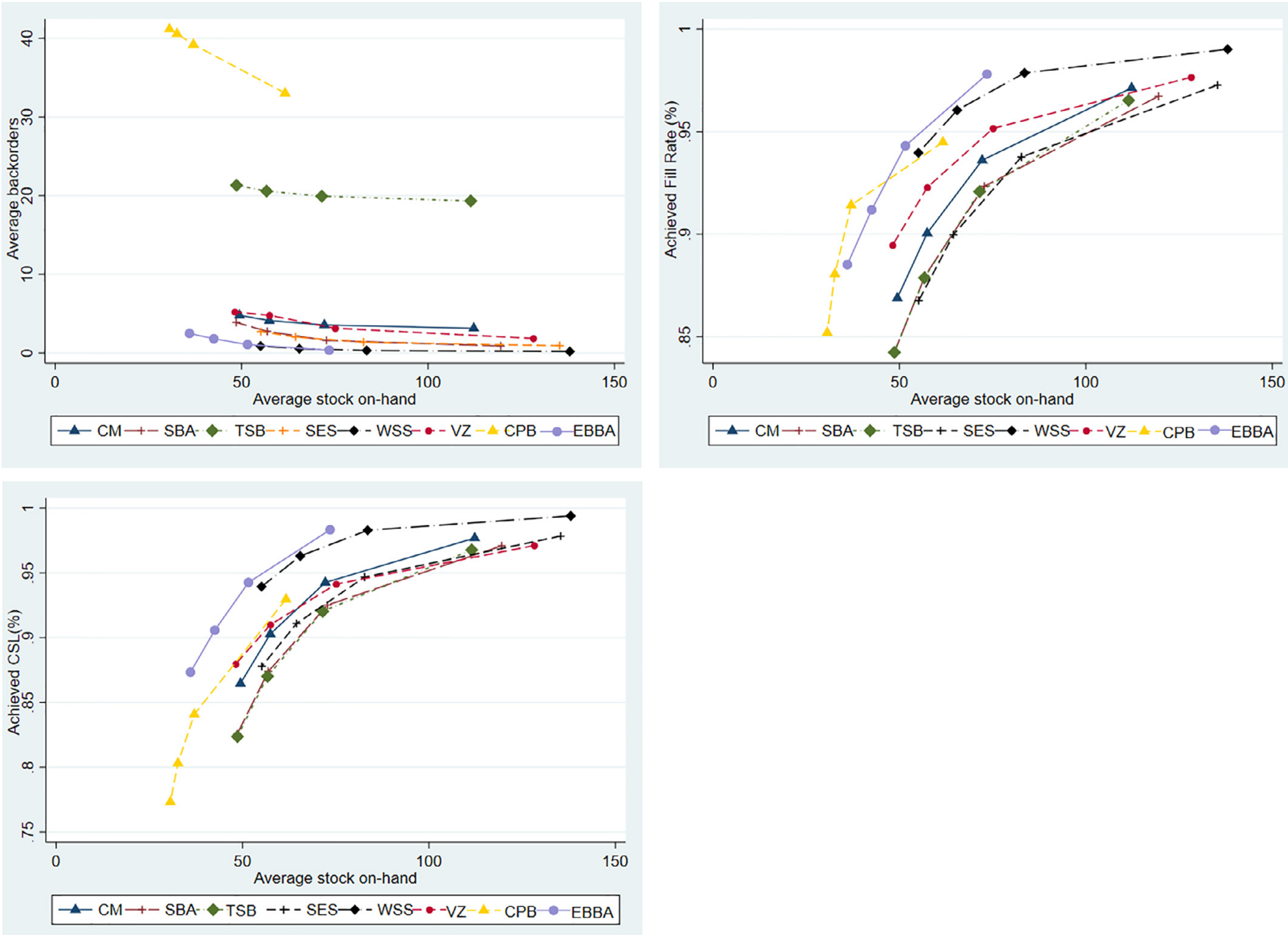

Fig. 1. (a). Stock on-hand versus backorders. (b). Stock on-hand versus FR. (c). Stock on-hand versus CSL.

80 percent of the total number of SKUs) into a subset denoted as $S^{EBBA}_{Superior}$. Similarly, the remaining 9272 SKUs with inferior performance of EBBA are included in subset $S^{EBBA}_{Inferior}$. We conduct the empirical forecast-inventory analysis on each SKU subset and report the results for $S^{EBBA}_{Superior}$ in Table 3 and $S^{EBBA}_{Inferior}$ in Appendix E, Table E2. For consistency with the earlier studies (e.g., Babai et al., 2021; Hasni et al., 2019; Syntetos et al., 2015), we report the mean values of the performance metrics for comparison.

Table 3 reveals that for $S^{EBBA}_{Superior}$, WSS remains the best performing method in terms of achieving the target CSLs, but is the worst performer in terms of average stock on-hand. Conversely, CPB requires the least average stock on-hand but is the worst performer in attaining target CSLs by a large margin. EBBA's performance for $S^{EBBA}_{Superior}$ is much improved over the full dataset. EBBA now exceeds two of the target CSLs and is within 1% of meeting two targets. Except for WSS, all the other methods either have similar or lower achieved CSLs in comparison with EBBA. Meanwhile, exception for CPB, EBBA requires substantially less average stock on-hand than the other competing methods. In terms of the number of backorders and fill rate customer service metrics, EBBA ranks second even though it requires substantially less inventory than the other methods with the exception of CPB.

In order to draw a more conclusive understanding of the forecasting methods' relative inventory efficiency, we constructed three tradeoff curves using Table 3's results Fig. 1a-1c compare the average inventory on hand versus the three customer service metrics: backorders, fill rate and CSL, respectively. In Fig. 1a, the lowest curve is the best, while the highest curve is best in Figs. 1b and 1c. These results indicate EBBA more efficiently deploys inventory in achieving higher levels of customer service across the three metrics. As such, the performance concerns suggested by Table 2 appear due to a relatively small proportion (20%) of SKUs in our dataset. This supports the notion that EBBA's flexibility to effectively modeling a wide variety of demand pattern characteristics supports its robustness in finding relatively high quality solutions. We also performed a similar analysis for each of the other seven methods, but none of them demonstrated substantial performance improvement in achieving target CSLs with low average on-hand inventory and backorder levels.

For completeness, we provide summary results for the analysis of the $S^{EBBA}_{Inferior}$ in Appendix E, Table E2, which indicates, as expected, substantially lower performance of EBBA in achieving target CSLs. To ensure proper application of our proposed method, we must answer a critical question: What demand pattern worsens EBBA's performance? In this regard, Syntetos, Boylan and Croston (2005) categorize demand patterns based on the average inter-demand interval (ADI) and the squared coefficient of variation of demand size CV^2. They suggest demarcation values of 1.32 and 0.49 for ADI andCV^2, respectively Fig. 2. presents the four demand categories with the number of SKUs from the full empirical dataset belonging to each of the categories Table 4. reports the proportion

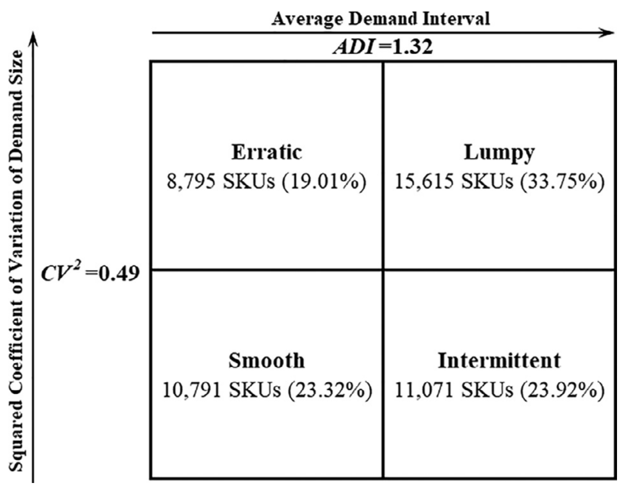

Fig. 2. SKU Demand Pattern Classification and Number of SKUs per Category.

of SKUs in each category belonging to the $S^{EBBA}_{Superior}$ and $S^{EBBA}_{Inferior}$ subsets, respectively.

The results suggest that EBBA performs best for intermittent SKUs and worst for erratic SKUs. But this demand pattern classification, based solely on ADI and CV^2, may not sufficiently explain the highly contrasting performance of EBBA shown in Tables 3 and E2. This is because, except for the intermittent category, in the other three categories, the proportion of SKUs belonging to the two subsets is not much different from the proportion of the two subsets in the full dataset.

Table 4
Proportion of SKUs Belonging to the Two Subsets.

Demand Category	$S^{EBBA}_{Superior}$	$S^{EBBA}_{Inferior}$
Erratic	5872(66.77%)	2923(33.23%)
Lumpy	12,485(79.96%)	3130(20.04%)
Smooth	8287(76.79%)	2504(23.21%)
Intermittent	10,356(93.54%)	715(6.46%)

Instead, the simulation results presented in Table E1 in Appendix E give a more convincing explanation: EBBA may exhibit inferior performance due to its underestimation of larger demand sizes caused by specific changes in a demand distribution over time. To confirm this answer, we evaluate relative changes over time of several measures of the demand distributions from the 56 in-sample periods to the 25 out-sample periods for each SKU in the two subsets. That includes the nonzero demand mean, standard deviation and CV^2 for measuring the demand distribution location and variation, skewness and Kurtosis for measuring the distribution shape, and ADI for measuring the distribution's inter-

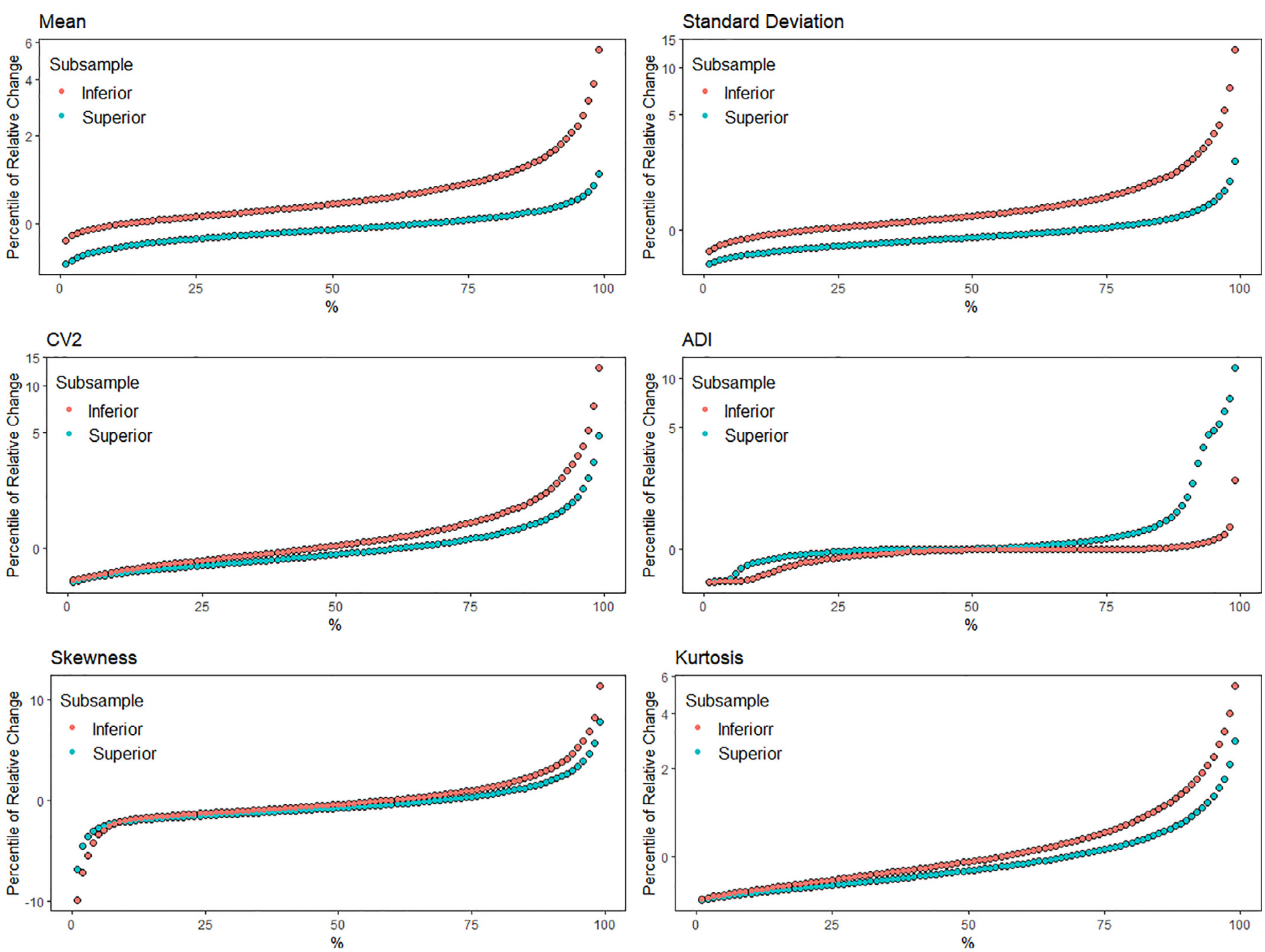

Fig. 3. Over Time Change in Demand Distribution from In-sample Periods to Out-Sample Periods.

mittency. The relative change is calculated as the difference of one measure's out-sample value and in-sample value as proportion of the in-sample value. For example, if one SKU has in-sample mean of 100 and out-sample mean of 150, its relative change in mean is 0.5, which means 50% increase in nonzero demand's mean from the in-sample periods to the out-sample periods.

Fig. 3 presents a comparison of SKUs in the subset $S^{EBBA}_{Superior}$ with those in the subset $S^{EBBA}_{Inferior}$ on the 1% to 99% percentiles of the relative changes in the different distribution measures.[2] We note that the relative changes in skewness and Kurtosis have similar percentile values between the two subsets. This implies the EBBA's performance difference between the two subsets is not notably caused by any particular distribution shape and its changes over time. For the distribution location and variation, the relative change in CV^2 displays a similar pattern between the two subsets that suggests the demand's standard deviation changes in proportion with its mean consistently for the SKUs across the two subsets. But it is shown that most SKUs in the subset $S^{EBBA}_{Inferior}$ have an upward shift in the demand's mean (also in standard deviation since it changes in proportion with the mean). Moreover, more than 50% of SKUs in this subset experience a relative change higher than 0.4 (i.e., increase by more than 40%) in the mean. By contrast, 80% of SKUs in the subset $S^{EBBA}_{Superior}$ have a relative change in the mean between −0.4 and 0.4. Only 5% of SKUs in this subset have a relative change higher than 0.4, while the other 15% show a large downward shift in the mean demand with a relative change lower than −0.4. Furthermore, there is a clear difference between the two subsets on the relative change in ADI: More than 20% of SKUs in the subset $S^{EBBA}_{Superior}$ have ADI increasing by higher than 40%; by contrast, only 2% of SKUs in the subset $S^{EBBA}_{Inferior}$ have an increase in ADI by higher than 40%.

In summary, our study shows EBBA is robust against a downward shift or a mild upward shift in the nonzero demand's mean (and standard deviation) over time and adapts quickly to a change in demand intermittency pattern (especially, a sharp increase in intermittency frequency), which align with our previous discussion of technical features of this method. Moreover, following our argument in Section 2 and demonstrated by the empirical results in this section, as a bootstrapping-based forecasting method EBBA can flexibly adjust to irregular shapes of nonzero demand distribution and its changes over time. On the other hand, one clear performance constraint for this method is a significant upward shift in the mean (and standard deviation) of the nonzero demand. In Appendix F, Figs. F1-F3 illustrate three typical SKUs from our dataset for which EBBA generally outperforms the other methods, and another SKU for which it performs inferiorly is illustrated in Fig. F4.

5. DISCUSSION and conclusion

Over the last two decades, the research on inventory forecasting has proposed many analytical methods for forecasting spare parts demand. Concurrently, intensive empirical studies have been conducted to evaluate and compare forecasting performance of these methods based on empirical industry data. These comparative studies often reach different conclusions on which method performs best that inevitably depend on the particular demand patterns reflected by the empirical datasets used in the studies. In consequence, the demand forecasting literature has a two-fold impact on inventory management practices. On the one hand, inventory managers have a rich forecasting toolkit from which they can select the best performing technique to model their demand data, assuming the data matches the demand pattern for which the model is best suited. On the other hand, new challenges always emerge in inventory control of spare parts. One particular challenge faced by our industrial collaborator is that the company is managing a large spare parts portfolio that displays highly complex demand patterns due to (a) strong interdependence of demand intermittency and (b) highly irregular demand distribution characterized by frequent but unexpected occurrence of certain demand sizes. Such demand distributions are not well governed by any standard probability distribution. We anticipate this challenge concerns a wide range of companies. Our study shows that popular forecasting methods may not provide a satisfying solution for this challenge. The currently available methods have limited flexibility, and each performs ideally for a narrow range of demand patterns. Parametric methods are constrained by their specific demand distribution assumptions. Although some distribution assumptions, such as compound Poisson distributions, fit spare parts demand data better than the others, none of them can satisfyingly deal with our scenario of strongly interdependent demand intermittency and an extremely wide range of different demand patterns. The existing nonparametric bootstrapping methods have been shown to perform relatively better for complex demand distribution. But those methods have their roots in Efron's frequentist bootstrapping approach and thus are essentially subject to a strong uninformative prior assumption that can significantly impact their performance when the demand distribution's complexity sharply increases.

To respond to this challenge, we develop a new nonparametric Bayesian approach to forecasting spare parts demand with its roots in the empirical Bayes paradigm. The new method (EBBA) is aimed to be a 'one for all' forecasting approach that is subject to few performance constraints and is highly flexible in dealing with extreme demand complexity caused by interdependence of demand intermittency and irregular demand distribution. EBBA has several attractive features. First, for each forecast item, EBBA resamples observed demand data based on their empirical probability density values enabling it to adjust to widely different demand patterns associated with different items and even accommodate nonstationary distribution to some extent. Secondly, EBBA models the mechanism of demand intermittency using a more flexible Bayesian approach so that it adapts quickly to a change in demand intermittency pattern. We assess the forecasting performance of EBBA against other popular nonparametric and parametric methods using an empirical demand dataset for 46,272 spare part SKUs from the heavy truck industry. The dataset includes items with a variety of challenging demand patterns. Our experimental results show that EBBA achieves the best overall forecast-inventory performance for a majority of the SKUs. It requires significantly less inventory on hand to satisfy a target customer service level, minimize backorders, and provide high inventory fill rates. The only performance weakness shown in our study occurs with a significant upward shift in the mean (and standard deviation) of the nonzero demand over time. Another relevant insight revealed by our study is that the dataset employed in our study can be used as a heavy stress test for the existing methods to check their flexibility in dealing with highly complex demand patterns. Such demand complexity is not explicitly reflected in the demand pattern classification suggested in Syntetos et al. (2005). While most of the existing methods included in our study tend to fall short in attaining target customer service levels, we note the superiority of WSS in meeting this goal in the presence of high complexity in demand pattern (but with a considerably larger inventory requirement.). As such, WSS may be an alternative to EBBA for SKUs with demand characteristics not well-suited for EBBA as identified in the paper.

[2] In order to enhance readability of Figure 3, we do not include the minimum (0% percentile) and maximum (100% percentile) relative changes in the figure.

It should be noted that EBBA inevitably requires more CPU than some simple methods such as SES. Also, our model structure is more complex than that of WSS which is already difficult for practitioners to understand (Syntetos et al., 2015). Furthermore, its theoretical foundations include advanced Bayesian concepts that can be challenging for many practitioners too. But this new method can be much simpler than the other methods in the challenge of a spare parts portfolio with highly complex demand patterns. In such a case, to properly use some conventional methods, managers must devote many efforts to developing a thorough understanding of the many different demand patterns in their managed inventory portfolio. By contrast, we should expect good overall forecasting results for this portfolio by using EBBA as long as it does not contain many items showing a significant upward shift in their nonzero demand mean and standard deviation over time. Boylan and Syntetos (2021) note that what constitutes "simple" and "complex" methods and how the best tradeoffs are achieved should be clearly revealed in intermittent demand forecasting. So, we suggest practitioners carefully weigh between the technical and theoretical complexity of this new method and the complexity of demand patterns appearing in inventory portfolios they are managing. If the latter clearly overweighs the former, we recommend using this new method to produce forecasts.

We suggest several promising future research directions. First, to the authors' best knowledge, our study is the first demonstration of the usefulness of empirical Bayes methods in forecasting spare parts demand. We hope this study stimulates future research efforts on developing forecasting methods in this category, which can be more flexible than both traditional frequentist and Bayesian methods. Second, we suggest more research efforts toward developing novel forecasting methods that can better deal with nonstationary demand distribution. The recent literature shows Bayesian model searching and averaging techniques have promising applications in this area (Hadj-Amar, Rand, Fiecas, Lévi & Huckstepp, 2020). Third, an effective combination of expert opinions and statistical forecasting can always improve forecast accuracy and inventory performance. But the real-world case motivating our work indicates that to formulate their expert opinions, practitioners need more advanced knowledge about managing high complexity in spare parts demand. This can also be a promising direction for future research in inventory management.

Acknowledgments

The authors are grateful to Editor Ruud Teunter and two anonymous reviewers for their constructive comments throughout the review process.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2022.02.033.

Appendix A: Mathematical description of Empirical Bayes Approach

Theoretically, as described in Carlin and Louis (2000), suppose an observed data sample $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]$ is generated from a distributional model $p(X|\theta)$ with the unknown parameter θ (which is $\boldsymbol{\pi}$ in our study). To obtain the Bayesian estimate of θ from $\boldsymbol{y}$, we use Bayes' theorem

$$p(\theta|\boldsymbol{y}) = p(\boldsymbol{y}|\theta)p(\theta)/f(\boldsymbol{y})$$

where $p(\theta|\boldsymbol{y})$ is the posterior distribution of θ given $\boldsymbol{y}$ and $p(\theta)$ is its prior distribution. Let $p(\theta)$ condition on a hyperparameter η (which is $\boldsymbol{\alpha}_n$ in our study), $p(\theta|\eta)$, and we have

$$p(\theta|\boldsymbol{y},\ \eta) = p(\boldsymbol{y}|\theta)p(\theta|\eta)/p(\boldsymbol{y}|\eta). \tag{A.1}$$

Then, we believe information about η is captured by the data distribution $p(\boldsymbol{y}|\eta)$ so that we can estimate η as a function of $\boldsymbol{y}$, $\hat{\eta} = \hat{\eta}(\boldsymbol{y})$. Consequently, we approximate the posterior distribution of $p(\theta|\boldsymbol{y}) \cong p(\theta|\boldsymbol{y},\ \hat{\eta})$, instead of marginalizing out η. As such, empirical Bayes can be considered as an approximation to a fully Bayesian treatment of a hierarchical model.

The empirical Bayes paradigm also has a technical justification. We alternatively write the equation (A.1) to be

$$p(\eta|\boldsymbol{y},\ \theta) = p(\boldsymbol{y}|\eta)p(\eta|\theta)/p(\boldsymbol{y}|\theta) \tag{A.2}$$

so that (A.1) and (A.2) form a conceptual structure for Gibbs sampling (Casella and George, 1992) in Bayesian data analysis: Choose arbitrary starting values $\theta = \theta^{(0)}$ and $\eta = \eta^{(0)}$ and initiate an iteration process:

Step 1: draw $\theta^{(1)} \sim p(\theta|\boldsymbol{y},\ \eta^{(0)})$

Step 2: draw $\eta^{(1)} \sim p(\eta|\boldsymbol{y},\ \theta^{(0)})$.

Then, do the process again but using $\theta^{(1)}$ and $\eta^{(1)}$ as new starting values to complete one iteration. After repeating this iteration sufficiently many times, we will obtain good approximations to $p(\theta|\boldsymbol{y})$ and $p(\eta|\boldsymbol{y})$. Empirical Bayes data analysis is qualitatively equivalent to partially executing this Gibbs sampler by using an estimate of η from observed data, $\hat{\eta} = \hat{\eta}(\boldsymbol{y})$, to approximate $p(\theta|\boldsymbol{y}) \cong p(\theta|\boldsymbol{y},\ \hat{\eta})$. So technically we may view empirical Bayes as frequentist intervention on a fully Bayesian data analysis.

Appendix B: Description of the other seven forecasting methods

y_t :Actual demand of an item in period t.

$\hat{y}_t$:Estimate of demand for an item in period t.

$\hat{z}_t$:Estimate of demand size for an item in period t.

$\hat{p}_t$:Estimate of demand occurrence indicator for an item in period t.

Method1: Croston's method (CM)

$$\hat{z}_t = \begin{cases} \hat{z}_{t-1} & if\ \ y_t = 0 \\ \hat{z}_{t-1} + \alpha(y_t - \hat{z}_{t-1}) & if\ \ y_t \neq 0 \end{cases}$$

$$\hat{p}_t = \begin{cases} \hat{p}_{t-1} & if\ \ y_t = 0 \\ \hat{p}_{t-1} + \beta(Q - \hat{p}_{t-1}) & if\ \ y_t \neq 0 \end{cases}$$

$$Q = \begin{cases} Q+1 & if\ \ y_t = 0 \\ 1 & if\ \ y_t \neq 0 \end{cases}$$

$$\hat{y}_t = \frac{\hat{z}_t}{\hat{p}_t}$$

Method 2: Syntetos-Boylan Approximation method (SBA)

$$\hat{z}_t = \begin{cases} \hat{z}_{t-1} & if\ \ y_t = 0 \\ \hat{z}_{t-1} + \alpha(y_t - \hat{z}_{t-1}) & if\ \ y_t \neq 0 \end{cases}$$

$$\hat{p}_t = \begin{cases} \hat{p}_{t-1} & if\ \ y_t = 0 \\ \hat{p}_{t-1} + \beta(Q - \hat{p}_{t-1}) & if\ \ y_t \neq 0 \end{cases}$$

$$Q = \begin{cases} Q+1 & if\ \ y_t = 0 \\ 1 & if\ \ y_t \neq 0 \end{cases}$$

$$\hat{y}_t = \left(1 - \frac{\alpha}{2}\right)\frac{\hat{z}_t}{\hat{p}_t}$$

Method 3: Teunter, Syntetos, and Babai's method (TSB)

$$\hat{z}_t = \begin{cases} \hat{z}_{t-1} & if\ \ y_t = 0 \\ \hat{z}_{t-1} + \alpha(y_t - \hat{z}_{t-1}) & if\ \ y_t \neq 0 \end{cases}$$

$$\hat{p}_t = \begin{cases} \hat{p}_{t-1} + \beta(0 - \hat{p}_{t-1}) & if\ \ y_t = 0 \\ \hat{p}_{t-1} + \beta(1 - \hat{p}_{t-1}) & if\ \ y_t \neq 0 \end{cases}$$

$$\hat{y}_t = \hat{z}_t\hat{p}_t$$

Method 4: Simple exponential smoothing method (SES)

$$\hat{y}_t = \alpha y_t + (1-\alpha)\hat{y}_{t-1}$$

Method 5: Willemain, Smart, and Schwarz's method (WSS)

Step1: Obtain historical demand data in chosen time buckets (e.g., days, weeks, months);

Step 2: Estimate transition probabilities for a two-state (zero vs. non-zero) Markov model;

Step 3: Conditional on last observed demand, use the Markov model to generate a sequence of zero/non-zero values over the forecast horizon;

Step 4: Replace every non-zero state marker with a numerical value sampled at random, with replacement, from the set of observed non-zero demands;

Step 5: Jitter the nonzero demand values.

Jittered = 1+ INT $\{X^*+Z\sqrt{X^*}\}$, if Jittered ≤ 0, then Jittered = X^*.

Step 6: Sum the forecast values over the horizon to get one predicted value of LTD.

Step 7: Repeat Step3–6 several times.

Step 8: Sort and use the resulting distribution of LTD values.

Method 6: Viswanathan and Zhou's method (VZ)

Step 1: Obtain histogram of the historical demand data(including both demand size data and demand interval data) in a chosen time bucket.

Step 2: Randomly generate demand interval according to the corresponding histogram. Update the time horizon, which is used to count the time passing by.

Step 3: If the time horizon is equal to or less than the leadtime, randomly generate demand size according to the demand size interval histogram.Then go to Step2. Else, sum the generated demand sizes over the leadtime and get one predicted value of the leadtime demand. Then go to Step4.

Step 4: Repeat Step 2–3 1000 times.

Step 5: Sort and use the resulting distribution of LTD values.

Method 7: Compound-Poisson Bayesian (CPB)

Actual demand y_t is assumed to follow a compound Poisson-Geometric distribution $PG(y_t|\lambda, \theta)$ in which demand arrivals have a Poisson distribution with the parameter λ and the demand sizes have a Geometric distribution with the parameter θ. In addition, the accumulation of actual demands over n periods is assumed to be a Lévy process so that the aggregate demand $T = \sum_{i=1}^{n} y_i$ has the distribution $PG(T|n\lambda, \theta)$. Meanwhile, λ is assigned the exponienal prior distribution, $f(\lambda) = e^{-\lambda},\ \lambda \in (0, \infty)$, and θ is assigned the uniform prior distribution $f(\theta) = 1,\ \theta \in [0, 1]$. As such, the prediction distribution of y_t, $P(y_t|T)$, is given by

$$P(y_t|T) \propto \iint PG(y_t|\lambda, \theta) PG(T|n\lambda, \theta) e^{-\lambda} d\lambda d\theta$$

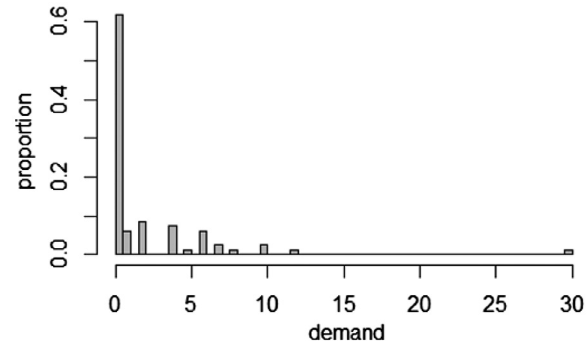

Fig. C1. Histogram of a randomly selected item.

Appendix C: Sample forecast output of EBBA

We illustrate the BBA's forecasting output for a randomly selected item from the empirical datasets Fig. C1. displays the item's histogram of bi-weekly demand Fig. C2. shows the corresponding distributional forecasts (posterior distributions) associated with replenishment cycle durations of one, two, three, and six bi-weekly time periods. The X-axis and Y-axis indicate the demand values and their occurrence frequencies, respectively. As illustrated, the posterior distributions emulate the characteristics of the empirical demand distributions, while smoothing demand randomness to arrive at the demand forecast for the replenishment cycle.

The distributional forecasts provide sufficient data for calculating the item's mean, standard deviation, and probability of zero demand during the replenishment cycle. Moreover, this approach enables estimating the target order-up-to inventory level to achieve a specified in stock service level. Assuming a periodic review (TR) order up to inventory model, Table C1 illustrates the type of managerial information that can be derived from the distributional forecast. The table data assumes a one-period review interval (T) and delivery lead time (L) of zero, one, two, or five bi-weekly time periods, respectively. For example, assuming a two-period replenishment cycle (i.e., $L = 1$ and $T = 1$), the predicted mean demand is 5.26 units, the standard deviation is 7.50, and the probability of zero demand is 0.35. The order-up-to level R for no stock out ser-

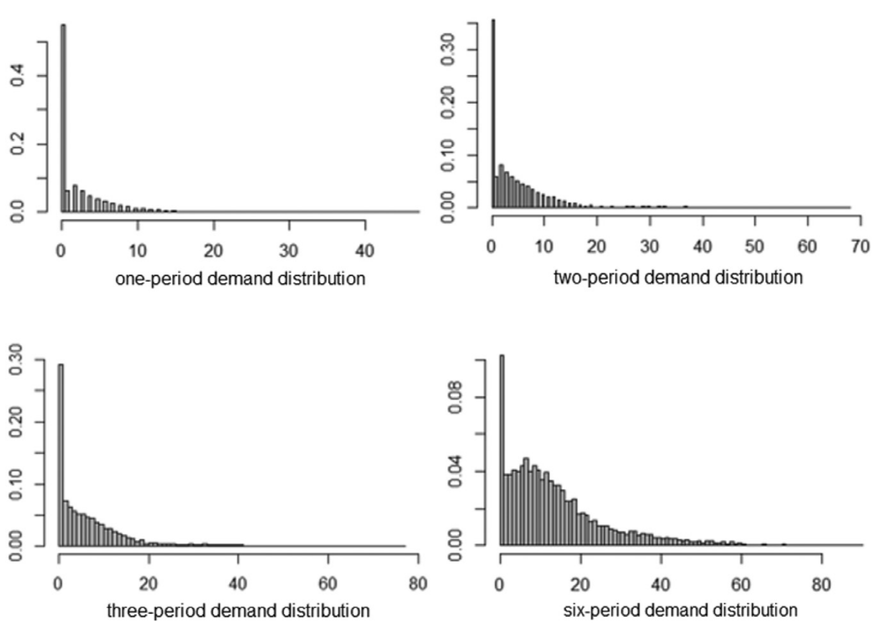

Fig. C2. Histogram of forecasted demand distributions.

Table C1
Illustrative Managerial Output Provided by EBBA for an SKU.

Lead Time	Review Interval	Mean	Standard deviation	Probability of zero demand	Order-up-to level based for varying Service Levels				
					50%	80%	90%	95%	99%
0 period	1 period	2.79	5.41	0.55	0	5	8	11	31
1 period	1 period	5.26	7.50	0.35	3	9	13	19	37
2 periods	1 period	7.68	9.21	0.23	5	12	18	29	42
5 periods	1 period	14.85	13.10	0.07	12	23	33	43	58

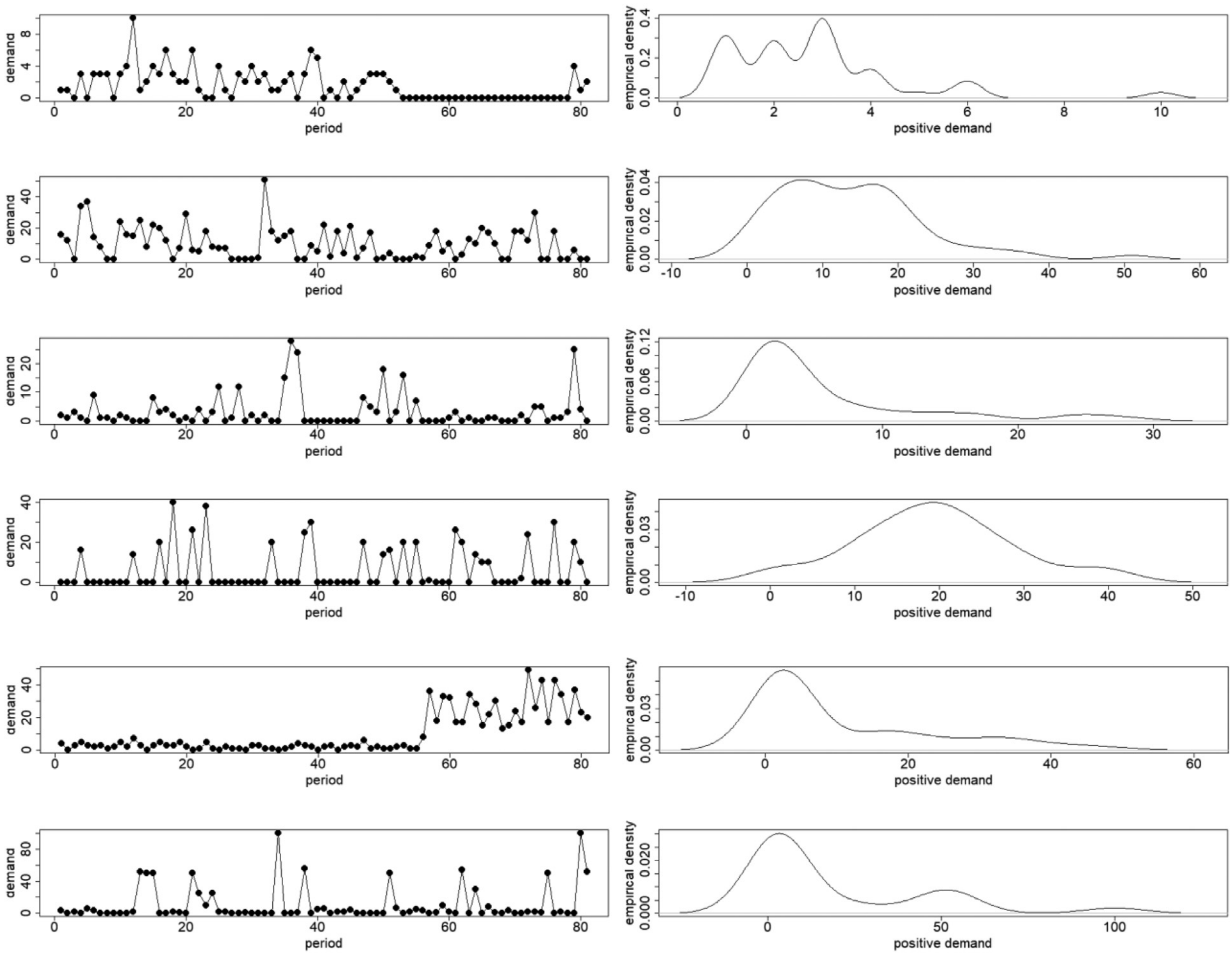

Fig. D1. Observed Biweekly Demand of Six Spare Parts and Kernel Approximations of Nonzero Demand.

vice levels of 90, 95 and 99 percent are 13, 19 and 37 units, respectively. The approach is equally applicable for reorder point and periodic review inventory systems with adjustments for replenishment cycle duration.

Appendix D: Summary Statistics and Illustrations of Demand Data Pattern in Our Empirical Dataset

.

Appendix E: Empirical Assessment of Performance of EBBA

Table E1
Distributional Accuracy of the RCD Pooled Sample.

Replenishment Cycle Length	Method	MIN	Estimated Percentiles of the RCD Pooled Sample								
			20%	40%	50%	60%	80%	90%	95%	99%	MAX
1 period	EBBA	0.00	0.22	0.35	0.47	0.57	0.83	0.95	0.98	1.00	1.00
	WSS	0.00	0.22	0.33	0.44	0.50	0.75	0.87	0.94	0.99	1.00
	VZ	0.00	0.30	0.40	0.50	0.63	0.85	0.94	0.97	1.00	1.00
	CPB	0.00	0.23	0.39	0.57	0.71	0.94	0.98	0.99	1.00	1.00
	Normal Distribution										

(*continued on next page*)

Table E1 (*continued*)

Replenishment Cycle Length	Method	MIN	Estimated Percentiles of the RCD Pooled Sample								
			20%	40%	50%	60%	80%	90%	95%	99%	MAX
	CM	0.00	0.24	0.34	0.40	0.47	0.71	0.89	0.97	1.00	1.00
	SBA	0.00	0.25	0.36	0.43	0.51	0.79	0.94	0.99	1.00	1.00
	TSB	0.00	0.24	0.34	0.40	0.47	0.72	0.90	0.98	1.00	1.00
	SES	0.00	0.25	0.35	0.41	0.47	0.72	0.89	0.97	1.00	1.00
	Negative Binomial Demand Distribution										
	CM	0.00	0.21	0.37	0.47	0.61	0.82	0.92	0.96	0.99	1.00
	SBA	0.00	0.23	0.40	0.51	0.65	0.86	0.94	0.98	1.00	1.00
	TSB	0.00	0.21	0.35	0.47	0.59	0.81	0.92	0.96	1.00	1.00
	SES	0.00	0.23	0.36	0.46	0.59	0.80	0.91	0.96	0.99	1.00
2 periods	EBBA	0.00	0.18	0.41	0.47	0.60	0.86	0.97	0.99	1.00	1.00
	WSS	0.00	0.16	0.34	0.44	0.51	0.75	0.89	0.96	1.00	1.00
	VZ	0.00	0.15	0.32	0.44	0.56	0.82	0.94	0.98	1.00	1.00
	CPB	0.00	0.10	0.20	0.27	0.39	0.83	0.97	0.99	1.00	1.00
	Normal Demand Distribution										
	CM	0.00	0.07	0.18	0.23	0.29	0.45	0.62	0.79	0.99	1.00
	SBA	0.00	0.08	0.19	0.24	0.30	0.48	0.68	0.86	1.00	1.00
	TSB	0.00	0.08	0.18	0.22	0.28	0.45	0.62	0.80	0.99	1.00
	SES	0.00	0.08	0.18	0.23	0.29	0.46	0.63	0.80	0.99	1.00
	Negative Binomial Demand Distribution										
	CM	0.00	0.03	0.14	0.21	0.29	0.57	0.78	0.89	0.98	1.00
	SBA	0.00	0.06	0.18	0.25	0.35	0.63	0.81	0.92	0.99	1.00
	TSB	0.00	0.05	0.15	0.20	0.27	0.56	0.77	0.88	0.99	1.00
	SES	0.00	0.06	0.15	0.21	0.32	0.55	0.75	0.87	0.98	1.00
3 periods	EBBA	0.00	0.14	0.38	0.50	0.64	0.91	0.98	1.00	1.00	1.00
	WSS	0.00	0.12	0.31	0.41	0.52	0.78	0.92	0.98	1.00	1.00
	VZ	0.00	0.11	0.31	0.42	0.53	0.82	0.95	0.99	1.00	1.00
	CPB	0.00	0.06	0.11	0.15	0.23	0.62	0.97	0.99	1.00	1.00
	Normal Demand Distribution										
	CM	0.00	0.02	0.10	0.15	0.20	0.34	0.47	0.65	0.96	1.00
	SBA	0.00	0.03	0.11	0.15	0.20	0.36	0.51	0.72	0.99	1.00
	TSB	0.00	0.02	0.10	0.15	0.20	0.34	0.52	0.70	0.97	1.00
	SES	0.00	0.02	0.10	0.15	0.20	0.38	0.53	0.72	0.97	1.00
	Negative Binomial Demand Distribution										
	CM	0.00	0.00	0.05	0.09	0.15	0.39	0.62	0.80	0.99	1.00
	SBA	0.00	0.01	0.07	0.12	0.19	0.45	0.68	0.84	0.99	1.00
	TSB	0.00	0.01	0.07	0.10	0.15	0.42	0.69	0.86	1.00	1.00
	SES	0.00	0.01	0.07	0.11	0.18	0.44	0.60	0.77	0.98	1.00

Table E2
Summary Experimental Results for the Subset with Inferior EBBA Performance (20% of SKUs).

Target Service Level 99%					Target Service Level 95%				
	Inventory on hand	Backorder	FR (%)	CSL (%)		Inventory on hand	Backorder	FR (%)	CSL (%)
CM	490.51	69.71	91.90%	92.12%	CM	247.89	77.04	87.01%	86.05%
SBA	549.77	44.92	91.53%	91.69%	SBA	254.38	54.87	85.78%	83.89%
TSB	500.60	200.32	90.53%	90.24%	TSB	242.39	208.42	84.90%	82.53%
SES	721.08	46.87	92.59%	92.75%	SES	284.71	56.24	86.49%	86.02%
WSS	741.76	12.86	93.91%	93.59%	WSS	191.41	31.54	88.02%	84.41%
VZ	720.93	24.26	90.77%	87.24%	VZ	173.31	51.88	84.34%	76.49%
CPB	111.46	278.13	82.90%	68.81%	CPB	85.92	303.19	72.28%	49.71%
EBBA	157.14	37.69	85.46%	80.96%	EBBA	109.88	61.92	77.36%	69.15%
Target Service Level 90%	Target Service Level 85%								
	Inventory on hand	Backorder	FR (%)	CSL (%)		Inventory on hand	Backorder	FR (%)	CSL (%)
CM	179.33	85.81	83.30%	80.10%	CM	144.71	94.06	80.48%	74.98%
SBA	174.07	69.34	81.41%	76.28%	SBA	136.39	82.78	78.11%	79.78%
TSB	169.96	218.05	80.73%	75.32%	TSB	133.98	227.04	77.54%	69.09%
SES	186.64	68.29	81.23%	79.59%	SES	147.55	79.41	77.81%	74.68%
WSS	141.59	46.05	82.84%	76.64%	WSS	116.45	57.34	79.08%	70.83%
VZ	125.52	71.28	79.06%	68.56%	VZ	105.57	81.66	75.38%	62.83%
CPB	82.47	307.82	68.40%	43.97%	CPB	81.01	309.76	66.13%	40.56%
EBBA	93.03	74.45	73.47%	62.08%	EBBA	84.67	83.55	70.96%	57.10%

Appendix F: Illustrations of Demand Distribution Change Over Time

Fig. F1, Fig. F2, Fig. F3, Fig. F4

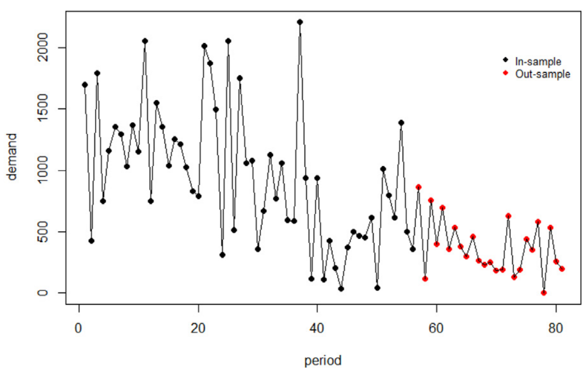

Fig. F1. Downward Shift in Nonzero Demand Mean and Standard Deviation.

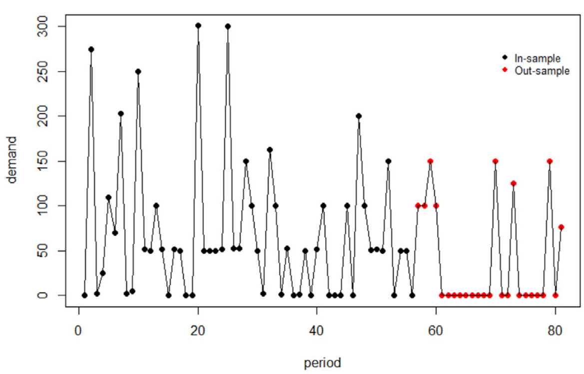

Fig. F2. Increase in Demand Intermittency with Strong Interdependence.

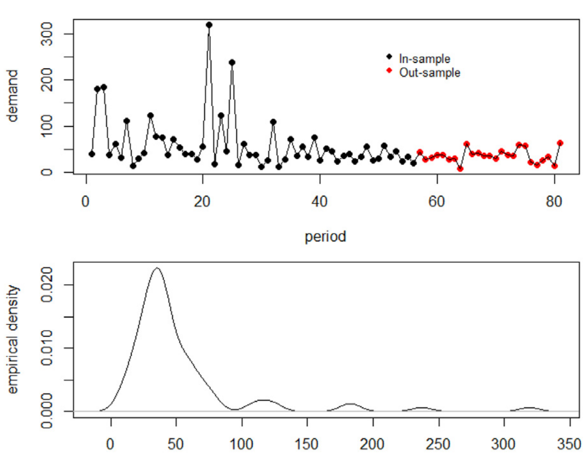

Fig. F3. Irregular Nonzero Demand Distribution with Extreme Tail Behavior.

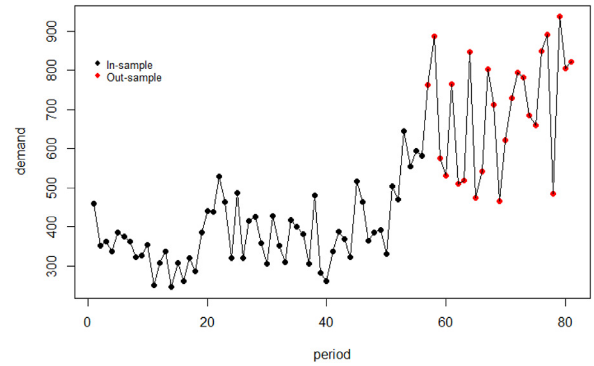

Fig. F4. Upward Shift in Nonzero Demand Mean and Standard Deviation.

.

References

Alfaro, M. E., Zoller, S., & Lutzoni, F. (2003). Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution, 20*(2), 255–266.

Aronis, K. P., Magou, I., Dekker, R., & Tagaras, G. (2004). Inventory control of spare parts using a Bayesian approach: A case study. *European journal of operational research, 154*(3), 730–739.

Assoudou, S., & Essebbar, B. (2004). A Bayesian model for binary Markov chains. *International Journal of Mathematics and Mathematical Sciences*, (8), 421–429 2004.

Babai, M. Z., Chen, H., Syntetos, A. A., & Lengu, D. (2021). A compound-Poisson Bayesian approach for spare parts inventory forecasting. *International Journal of Production Economics, 232*, Article 107954.

Babai, M. Z., & Syntetos, A. A. (2007). An empirical comparison of parametric and nonparametric forecasting methods for intermittent demand items. In *Proceedings of the OR49, Edinburgh* (pp. 4–6).

Babai, M. Z., Tsadiras, A., & Papadopoulos, C. (2020). On the empirical performance of some new neural network methods for forecasting intermittent demand. *International IMA Journal of Management Mathematics, 31*, 281–305.

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega, 40*, 722–737.

Boone, C. A., Craighead, C. W., & Hanna, J. B. (2008). Critical challenges of inventory management in service parts supply: A Delph study. *Operations Management Research, 1*(1), 31–39.

Boylan, J. E., & Syntetos, A. A. (2021). *Intermittent demand forecasting: Context, methods and applications*. New York: John Wiley & Sons.

Boylan, J. E., & Syntetos, A. A. (2010). Spare parts management: A review of forecasting research and extensions. *IMA Journal of Management Mathematics, 21*, 227–237.

Carlin, B. P., & Louis, T. A. (2000). Empirical Bayes: Past, present and future. *Journal of the American Statistical Association, 95*(452), 1286–1289.

Casella, G. (1992). Illustrating empirical Bayes methods. *Chemometrics and Intelligent Laboratory Systems, 16*(2), 107–125.

Cattani, K. D., Jacobs, F. R., & Schoenfelder, J. (2011). Common inventory modeling collaboration that fall short: Arborescent networks, Poisson demand, and single echelon approximations. *Journal of Operations Management, 29*(5), 488–499.

Chun, Y. H. (2016). Designing repetitive screening procedures with imperfect inspections: An empirical Bayes approach. *European Journal of Operational Research, 253*(3), 639–647.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly, 23*, 289–303.

Deloitte Consulting (2013). Driving aftermarket value: Upgrade spare parts supply chain, Deloitte China auto industry spare parts management benchmark survey, Deloitte White Paper. Available at https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/manufacturing/deloitte-cn-mfg-auto-indspareparts-whitepaper-en-171013.pdf.

Desomer, E. (2011). The service revolution: Building a world class service and parts business, Deloitte White Paper. Available at https://www2.deloitte.com/content/dam/Deloitte/cn/Documents/manufacturing/deloitte-cn-mfg-auto-indspareparts-whitepaper-en-171013.pdf.

Dolgui, A., & Pashkevich, M. (2008). Demand forecasting for multiple slow-moving items with short requests history and unequal demand variance. *International Journal of Production Economics, 112*, 885–894.

Eckert, F., Hyndman, R. J., & Panagiotelis, A. (2021). Forecasting Swiss exports using Bayesian forecast reconciliation. *European Journal of Operational Research, 291*(2), 693–710.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics, 7*(1), 1–26.

Forecast Pro, (2020). Available at https://www.forecastpro.com/solutions/forecast-pro/forecasting-methods/, accesses May 29, 2020.

Frazier, D. T., Robert, C. P., & Rousseau, J. (2020). Model misspecification in approximate Bayesian computation: Consequences and diagnostics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82*(2), 421–444.
Gardner, E., & Koehler, A. B. (2005). Comments on a patented bootstrapping method for forecasting intermittent demand. *International Journal of Forecasting, 21*(3), 617–618.
Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2021). Inventory - forecasting: Mind the gap. *European Journal of Operational Research*. https://doi.org/10.1016/j.ejor.2021.07.040.
Gönen, M., Johnson, W. O., Lu, Y., & Westfall, P. H. (2019). Comparing objective and subjective Bayes factors for the two-sample comparison: The classification theorem in action. *The American Statistician, 73*(1), 22–31.
Guijarro, E., Cardos, M., & Babiloni, E. (2012). On the exact calculation of the fill rate in a periodic review inventory policy under discrete demand patterns. *European Journal of Operational Research, 218*, 442–447.
Gutierrez, R. S., Solis, A. O., & Mukhopadhyay, S. (2008). Lumpy demand forecasting using neural networks. *International journal of production economics, 111*(2), 409–420.
Hadj-Amar, B., Rand, B. F., Fiecas, M., Lévi, F., & Huckstepp, R. (2020). Bayesian model search for nonstationary periodic time series. *Journal of the American Statistical Association, 115*(531), 1320–1335.
Hasni, M., Aguir, M. S., Babai, M. Z., & Jemai, Z. (2019). Spare parts demand forecasting: A review on bootstrapping methods. *International Journal of Production Research, 57*(15–16), 1–14.
Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
Hill, R. M. (1997). Applying Bayesian methodology with a uniform prior to the single period inventory model. *European Journal of Operational Research, 98*, 556–562.
Hong, L., & Martin, R. (2020). Model misspecification, Bayesian versus credibility estimation, and Gibbs posteriors. *Scandinavian Actuarial Journal*, (7), 634–649 2020.
Hu, Q., Chakhar, S., Siraj, S., & Labib, A. (2018). OR in spare parts management: A review. *European Journal of Operational Research, 262*, 1136–1163.
Jasper, J.B. (.2006). Quick response solutions FedEx critical inventory logistics, Revitalized White Paper. Available at https://docplayer.net/18237716-Quick-response-solutions.html
Kourentzes, N. (2013). Intermittent demand forecasts with neural networks. *International Journal of Production Economics, 143*(1), 198–206.
Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., & Gucci, S. (2017). Single-hidden layer neural networks for forecasting intermittent demand. *International Journal of Production Economics, 183*, 116–128.
Munyangabo, P., Waititu, A., & Wanjoya, A. K. (2019). Nested error nonparametric unit level model performance in the context of empirical Bayes (EB) approach. *Journal of Statistical and Econometric Methods, 8*(3), 35–48.
Nambiar, M., Simchi-Levi, D., & Wang, H. (2019). Dynamic learning and pricing with model misspecification. *Management Science, 65*(11), 4980–5000.
Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). Horses for Courses' in demand forecasting. *European Journal of Operational Research, 237*(1), 152–163.
Pince, C., Turrini, L., & Meissner, J. (2021). Intermittent demand forecasting for spare parts: A critical Review. *Omega*, 105.
Quigley, J., Hardman, G., Bedford, T., & Walls, L. (2011). Merging expert and empirical data for rare event frequency estimation: Pool homogenisation for empirical Bayes models. *Reliability Engineering & System Safety, 96*(6), 687–695.
Quigley, J., Walls, L., Demirel, G., MacCarthy, B. L., & Parsa, M. (2018). Supplier quality improvement: The value of information under uncertainty. *European journal of operational research, 264*(3), 932–947.
Rego, J. R., & Mesquita, M. A. (2015). Demand forecasting and inventory control: A simulation study on automotive spare parts. *International Journal of Production Economics, 161*, 1–16.
Robbins, H. (1955). An Empirical Bayes Approach to Statistics. In *Proceedings of the Third Berkeley Symposium Mathematical Statistics and Probability 1* (pp. 157–164). Berkeley: University of California Press.
Robert, C., & Casella, G. (2013). *Monte carlo statistical methods*. Springer Science & Business Media.
Rubin, D. B. (1981). The bayesian bootstrap. *The annals of statistics, 9*(1), 130–134.
Silver, E. A. (1965). Bayesian determination of the reorder point of a slow moving item. *Operations Research, 13*(6), 989–997.
Syntetos, A., Lengu, D., & Babai, M. Z. (2013). A note on the demand distributions of spare parts. *International Journal of Production Research, 51*(21), 6356–6358.
Syntetos, A. A., Babai, M. Z., & Altay, E. N. (2012). On the demand distributions of spare parts. *International Journal of Production Research, 50*(8), 2101–2117.
Syntetos, A. A., Babai, M. Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, P. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research, 252*, 1–26.
Syntetos, A. A., Babai, M. Z., & Gardner, E. S. (2015). Forecasting intermittent inventory demands: Simple parametric methods vs. bootstrapping. *Journal of Business Research, 68*, 1746–1752.
Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics, 71*, 457–466.
Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting, 21*, 303–314.
Syntetos, A. A., & Boylan, J. E. (2006). On the stock control performance of intermittent demand estimators. *International Journal of Production Economics, 103*(1), 36–47.
Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the Categorization of Demand Patterns. *The Journal of the Operational Research Society, 56*(5), 495–503.
Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research, 214*, 606–615.
Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2017). Stock keeping unit fill rate specification. *European Journal of Operational Research, 250*, 917–925.
Turrini, L., & Meissner, J. (2019). Spare parts inventory management: New evidence from distribution fitting. *European Journal of Operational Research, 273*, 118–130.
Viswanathan, S., & Zhou, C. X. (2008). *A new bootstrapping based method for forecasting and safety stock determination for intermittent demand items. working paper*. Singapore: Nanyang Business School, Nanyang Technological University.
Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20*(3), 375–387.
Wouters, S., & Rustenburg, J.W. (.2014). Spare parts management and optimisation—Pragmatism by differentiation, Gordian Logistic Experts White Paper. Available at https://gordian.nl/wp-content/uploads/2018/01/White-paper-Spare-Parts-Management-Optimisation-v1.2.pdf
Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*, 35–62.
Zhou, C., & Viswanathan, S. (2011). Comparison of a new bootstrapping method with parametric approaches for safety stock determination in service parts inventory systems. *International Journal Production Economics, 133*, 481–485.