



Invited Review

Inventory – forecasting: Mind the gap

Thanos E. Goltzos^a, Aris A. Syntetos^{a,*}, Christoph H. Glock^b, George Ioannou^c^a PARC Institute of Manufacturing, Logistics and Inventory, Cardiff Business School, Cardiff University, Aberconway Building, Colum Road, Cardiff CF10 3EU, United Kingdom^b Institute of Production and Supply Chain Management, Department of Law and Economics, Technical University of Darmstadt, Hochschulstraße 1, Darmstadt 64289, Germany^c Management Science Laboratory, Department of Management Science and Technology, Athens University of Economics and Business, Evelpidon 47A & Lefkados 33, 9th floor, Athens GR-113 62, Greece

ARTICLE INFO

Article history:

Received 10 February 2020

Accepted 21 July 2021

Available online 14 August 2021

Keywords:

Forecasting

Inventory control

Inventory forecasting

Literature review

ABSTRACT

We are concerned with the interaction and integration between demand forecasting and inventory control, in the context of supply chain operations. The majority of the literature is fragmented. Forecasting research more often than not assumes forecasting to be an end in itself, disregarding any subsequent stages of computation that are needed to transform forecasts into replenishment decisions. Conversely, most contributions in inventory theory assume that demand (and its parameters) are known, in effect disregarding any preceding stages of computation. Explicit recognition of these shortcomings is an important step towards more realistic theoretical developments, but still not particularly helpful unless they are somehow addressed. Even then, forecasts often constitute exogenous variables that serially feed into a stock control model. Finally, there is a small but growing stream of research that is explicitly built around jointly tackling the inventory forecasting question.

We introduce a framework to define four levels of integration: from disregarding, to acknowledging, to partly addressing, to fully understanding the interactions. Focusing on the last two, we conduct a structured review of relevant (integrated) academic contributions in the area of forecasting and inventory control and argue for their classification with regard to integration. We show that the development from one level to another is in many cases chronological in order, but also associated with specific schools of thought. We also argue that although movement from one level to another adds realism, it also adds complexity in terms of actual implementations, and thus a trade-off exists. The article makes a contribution into an area that has always been fragmented despite the importance of bringing the forecasting and inventory communities together to solve problems of common interest. We close with an indicative agenda for further research and a call for more theoretical contributions, but also more work that would help to expand the empirical knowledge base in this area.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

1. Motivation and background

Inventory control is concerned with supporting operational decisions on when and how much to replenish for each of multiple stock keeping units (SKUs), as well as the parts and materials used to make them. These inventories are in place to satisfy customer demand, at a required service level and/or a budget. In situations where demand is dependent (such as for parts and components in higher than 0 levels in the bill of materials), controlling the inventories boils down to a scheduling exercise through materials re-

quirement planning (MRP) procedures. If demand is independent, but somehow is known in advance, models that in their most basic version minimize the sum of (expected) ordering and inventory carrying costs (such as Harris', 1913, EOQ model or Wagner and Whitin's, 1958, model), and that also take account of constraints that have to be satisfied in a given situation (e.g., a minimum service level), are used. However, customer demand is typically independent and unknown at the time stocking and production decisions need to be made, and therefore we need to forecast it.

In this context, we refer to a forecast as the (best possible) genuine expectation of how much demand is going to be for a particular SKU (often with sales as a proxy)¹. Most often this refers

* Corresponding author.

E-mail addresses: goltzos@cardiff.ac.uk (T.E. Goltzos), syntetos@cardiff.ac.uk (A.A. Syntetos), glock@pscm.tu-darmstadt.de (C.H. Glock), ioannou@aub.gr (G. Ioannou).¹ The terms 'demand' and 'sales' are frequently used interchangeably in the forecasting and inventory control literature. However, while demand describes exactly

to point, mean demand forecasts, although forecasts of variance or higher moments, other quantiles or indeed the entire lead time demand distribution may be required. For the purposes of this article, we distinguish such forecasts of demand, to be used for SKU inventory management, from forecasts for other functions (e.g., marketing). We use the term “inventory forecasting” then to describe the intersection of these two areas, i.e. integrated literature of forecasting and inventory control. These works are manipulating characteristics of demand (or indeed of forecasts), in pursuit of inventory and ultimately supply chain efficiencies.

1.1. Background

The first articles on inventory control date back to the early 20th century. The to this day relevant, and aptly named article “How many parts to make at once” by Harris (1913) introduced the EOQ model. Basic EOQ formulations are built on the assumption that demand and its (true) parameters are known and constant (in a sense, a “perfect forecast” is available), with some exceptions (see Glock et al., 2014; Andriolo et al., 2014; for reviews of the EOQ literature). This thinking, that completely bypasses any need to forecast, is reflected also in several seminal inventory textbooks, and is not constrained to EOQ formulations. For example, forecasting is absent in Hadley and Whitin (1963) and Arrow et al. (1958), but also in more recent textbooks (Zipkin, 2000; Muckstadt and Sapra, 2010), Silver et al. (1998, 2017) and Axsäter (2015) take a step forward in explicitly recognising the need to forecast and dedicate a chapter to it.

Similarly, classical forecasting textbooks are not contextualised, treating forecasting as an end in itself (e.g., Makridakis et al., 1998; Ord et al., 2017; Hyndman and Athanasopoulos, 2018), with forecasting for inventory control being repeatedly reported a neglected area (e.g., Fildes and Beard, 1992; Prak et al., 2017). That is to say, they do not take into account what the purpose of the forecasts is (i.e. the forecast utility, be it in budgeting, energy, scheduling, or in our case, inventory control). Broadly speaking, in terms of forecasting performance, the forecasting literature so far has been mostly concerned with achieving gains against some point forecast error metrics - when, most often, forecasting the mean.

The implicit assumption here is that “achievable improvements in accuracy lead directly to worthwhile savings” (Fildes and Beard, 1992, pp. 24). An alarming number of works have, however, challenged this otherwise intuitive and common-sense conjecture (e.g., Flores et al., 1993; Eaves and Kingsman, 2004; Syntetos et al., 2010; Tratar, 2010; Babai et al., 2019; Kourentzes et al., 2020). Forecasting accuracy metrics can take a variety of forms, but especially when constrained to assessing the accuracy of point forecasts can fail to evaluate their impact (utility) to inventory control. As Davydenko and Fildes (2013, pp. 511) put it, “the key issue when evaluating a forecasting process is the improvements achieved in supply chain performance”, viz., the implications of any attained accuracy.

Inventory control performance, on the other hand, has been mostly concerned with attaining inventory efficiencies, often reported through (the trade-off between) inventory-related costs and some service achievement. However, in the majority of cases, this is done while either disregarding forecasting or assuming some idealistic forecast is available (conforming to strict, often unrealistic specifications).

what a customer would want to buy, sales represent what the customers did buy (and therefore lead to censored demand information; Conrad, 1976). In a wholesaling, B2B (business to business) or online sales environment, the real demand can be traced and documented (even when not met). In retail situations however, there is often no way to know what the real demand is as companies would tend to only document sales, which are then used as a proxy for demand, sometimes with some interpolation (see, e.g., Lau and Lau, 1996, and Tan and Karabati, 2004, for a review on the estimation of demand distribution based on censored sales data).

Historically, these two parts of the same “inventory forecasting” function have been treated as separate entities. Of course, this is by no means a critique to these works and authors. There is little doubt that they are important contributions to the state-of-knowledge at the time that they were written, and that the research they contain is important and relevant. Conceptually, these works provide the foundations for integration, defining the constituent blocks of inventory forecasting. Understanding the parts (forecasting and inventory control) is required before attempting to look at the whole (integrated inventory control and forecasting).

This isolationist approach is also reflective of the respective communities and conferences. The International Symposium on Inventories (ISIR) introduced an inventory forecasting stream only in 2008. Supply chain (and therein inventory) related streams were not popular in the International Symposium on Forecasting (ISF) until recently. Simple analysis of the International Journal of Production Economics (IJPE), a journal focused on production and operations management (and publishing research from the ISIR), and the International Journal of Forecasting (IJF) published on behalf of the International Institute of Forecasters (organisers of ISF) is telling (see Figure 1).

1.2. The need for integration

It has been shown that taking this isolationist approach is not always the best course in terms of performance. For example, and quite tellingly, the best demand forecasting method for minimising inventory costs is not necessarily always the one with the best forecast accuracy (Tratar, 2010; Kourentzes et al., 2020). In some instances, potential costly undershoots, caused by problematic inventory control assumptions with regards to the distribution of the forecast errors, are recouped by positive bias in the forecasts (e.g., Babai et al., 2014). These are irrefutably valid observations, however the inference that forecast bias may improve the system’s performance is problematic, and indicative of frail assumptions (Taylor, 2007; Syntetos and Boylan, 2008).

Further, even when the forecast accuracy and inventory performance improvements are in the same direction, they may be of different magnitude; Syntetos et al. (2010) found a 1% reduction in forecast accuracy to translate into a 10–15% reduction in inventory costs for comparable service levels, casting further doubt as to what extent accuracy measures may help explain inventory performance. The assumption then that forecast accuracy gains will translate into inventory gains does not appear to universally hold and seems contingent to the validity of further assumptions in subsequent inventory control calculations.

Simulations, and in particular ones with empirical data, have been extremely helpful in revealing the shortcomings of common-place inventory-related theoretical assumptions (e.g., Bretschneider, 1986; Eppen and Martin, 1988). This is also supported a fortiori by the fact that there is no consensus (yet) on ways to select a forecast, a persistent issue within the forecasting community (Gardner, 1985, 2006; De Gooijer and Hyndman, 2006; Kolassa, 2016).

What the above highlight is our limited understanding of the interrelations between forecasting and inventory control, in particular when operating far away from theoretical assumptions. Critically, it also highlights a frequent elusion to accommodate the fact that demand is actually forecasted rather than known (Prak et al., 2017). Even when the theoretical assumptions might be reasonable, the frequency at which judgemental interventions occur in practice (Fildes et al., 2009) either on forecasts (Trapero et al., 2011) or directly on inventory quantities of interest (e.g., re-order levels and

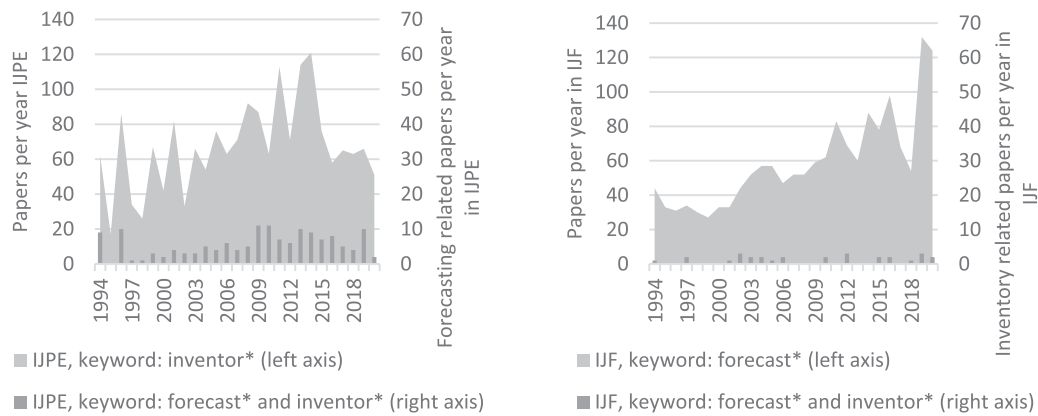


Fig. 1. In IJPE, left (IJF, right), just 197 (31) articles contain the words ‘inventor*’ and ‘forecast*’, out of 1952 (1735) containing just the word ‘inventor*’ (‘forecast*’), about 10% (2%). Notice the scale difference between the primary (left, area) and secondary (right, bar chart) axes in both figures. Source: Scopus, search in titles, abstracts, and keywords, up to 2020.

order quantities; Syntetos et al., 2016) might cast doubt on their robustness².

This is not a critique; it is rather a reminder of how open this area is to contributions. However, this is central to the argument for integration: there are intricate inventory forecasting problems that require to be approached as a system, taking into account that inventory decisions should be/are³ informed by forecasts. Control theory lends a nice structure of analysis (see Mason-Jones and Towill, 1998): In inventory management, we attempt to control demand uncertainty to efficiently meet customer demand. To do so, we introduce ‘control mechanisms’ in terms of forecasting procedures (feedforward control) and inventory policies (feedback control) (Towill, 1982). When the interactions of these ‘control mechanisms’ are not carefully explored, often unwarranted ‘control uncertainty’ is introduced (see, e.g., Goltso et al., 2019a).

One example is the bullwhip effect in a supply chain context, where (as we move further away from the customer in a supply chain) inventory oscillations become increasing multiples of end demand oscillations, inflating inventory costs (see, e.g., Li et al., 2014). Another example is the ‘issue point bias’ in an intermittent demand⁴ context, where a non-frequent demand occurrence drops the inventory level as it also inflates inappropriate forecasts of demand, triggering inflated orders that lead to overstocking (Croston, 1972).

We are of course not the first to point out the need to jointly approach the interrelated functions of forecasting and inventory control. For forecasting integration means that at a minimum, demand parameters used for inventory control need to be appropriately estimated and updated (Eppen and Martin, 1988), and that forecasting performance needs to be judged through inventory performance (Gardner, 1990). It is to use demand information to calculate (forecast) inventory quantities of interest, to take a holistic view to inventory forecasting solutions for a joint end goal. Watson (1987, pp. 82) observed that the use of an “elaborate reordering formula is rather pointless when demand-forecast fluctuations may

be large”, meaning that the elaboration should be expended on the demand forecast side. Gardner (pp. 498, 1990) noted that the forecasting aim should be “to improve customer service and reduce inventory investment”, meaning that forecasts should be judged by bottom-line inventory performance. A keynote speech in the practitioner stream of the International Symposium on Forecasting (ISF) in 2016 (Syntetos, 2016) was one vocal example of an increasing number of recent calls for more integrated approaches.

1.3. Summary and outline

This paper is in response to these calls for integration, an effort to understand and support integrated inventory forecasting. We attempt to consolidate relevant arguments, to serve as a single point of reference for researchers to address issues of mutual interest in the two communities. In order to do that we need to qualify the meaning of integration before we attempt to explore it. We explore what does and what does not constitute integration between the fields of forecasting and inventory control. At the same time, it is not our aim to criticise, nor do we imply any relationship between an article’s integration ‘level’ and the quality of research within.

The following notation is used. The letters I and F denote the focus of the paper, being inventory control and/or forecasting, respectively. The numbers zero to three indicate the level of integration. For papers whose focus is on inventory control, integration relates to the extent the work considers the fact that demand needs to be/is forecasted. For papers whose focus is on forecasting, integration relates to the extent the work actually considers that the ultimate goal is to achieve some inventory-related bottom line performance improvement (most commonly some service level/cost balance). Letters IF indicate integrated inventory forecasting literature. Figure 2 summarizes the framework:

- Level 0: Inventory application with **no mention** of forecasting (I0) or the inverse (F0)
- Level 1: Inventory application with **discussion** of forecasting (I1) or the inverse (F1)
- Level 2: **(Serial) application** of forecasting and inventory control (I2 or F2 according to focus)
- Level 3: **Integrated development** of inventory forecasting research (IF3)

We strive to adopt an objective position and avoid personal methodological preferences. We take a wide-lens snapshot of the literature in an attempt to qualify and quantify the degrees of integration between forecasting and inventory control. To do so, we perform a wide structured literature survey. To identify the keyword sets and establish our final sample, we construct and analyse

² Robustness here and later refers to the ‘sufficiently good’ performance of the policy under varying demand and product characteristics (Boylan and Syntetos, 2021; e.g., Arrow et al., 1958; Bijvank et al., 2014; for the backorder and lost sales cases of an order-up-to policy, respectively).

³ The be/is differentiation is important. Demand needs to be forecasted, and yet when demand is forecasted, the implications of whatever forecasting method is employed still need to be carefully considered (with relation to assumptions of the inventory model, see, e.g. Hsieh et al., 2020).

⁴ Infrequent positive demand interspersed with periods of zero demand. For the purposes of this work we also interchangeably discuss these demand patterns as ‘slow’. A thorough overview of intermittent demand forecasting and inventory control can be found in Boylan and Syntetos (2021).



Fig. 2. Integration levels.

a pre-sample of papers as well as contact various experts in the respective fields for suggestions.

There exists a very high degree of isolation between the two disciplines. When compared to the great body of research papers that constitute the forecasting and inventory control literatures, only a small fraction of papers are integrated (levels 2 and 3). This is illustrated in Section 2 using our keyword sets, but we note that this acts as a motivation rather than a hypothesis we set out to prove. We find that the first integrated approaches started appearing in the 1970s and have grown to about ten papers per year recently. We identify tracks of literature where integration is more prevalent, and report on various modelling decisions in this area.

The remainder of our paper is organised as follows: Section 2 describes the keyword selection process and reports on the survey protocol that was followed. This should facilitate the reproduction of our findings and constitute a starting point for further investigations in this area. Section 2 details the integration classification framework and the sample's classification flowchart. The classification framework is applied on our sample, and summary results are presented in Section 3. Main areas of interest for integration identified in Section 3 are isolated and further explored and discussed in Section 4. Section 5 summarises and attempts to recast the argument for integration, and discusses when it is warranted alongside promising pathways to pursue it.

2. Classification and review protocol

In this section we overview the classification process, the selection of our keyword sets and the compilation of our sample. It is not the aim of this paper to discuss the forecasting and inventory theory literature in its entirety. We are interested in the interaction and integration of forecasting and stock control. To this end, we have constructed our keywords sets to accordingly try and exclude inventory control articles that do not include forecasting elements and vice versa. An intentional and direct outcome of this is that non-integrated articles are underrepresented in our sample. All searches were conducted in Scopus, due to the breadth of databases it has access to.

2.1. Integration framework and classification

We introduce a classification framework, categorising papers by their level of integration and focus. It consists of four integration levels, alongside a designation for its focus (inventory or forecasting literature). The framework matured over a long period and has been presented and discussed numerous times in conferences and other events⁵. Feedback from the community and internal discussions led to numerous improvements and reclassifications. We are grateful to both the forecasting and inventory control communities

⁵ This work has been carried over five years and has over this time been presented to an International Institute of Forecasters workshop in Lancaster (2016), a research seminar at the Technical University of Darmstadt (2016), at the International Symposium on Inventories Research in Budapest (2018), and as a keynote speech (practitioner stream) for the International Symposium on Forecasting in Santander (2016).

for their valuable feedback that has led to this final version of our integration framework.

On one end (level 0), we find inventory applications disconnected from forecasting or the inverse. Inventory research here adopts convenient demand assumptions (cancelling the need to forecast, e.g., Zipkin, 2000; Muckstadt and Saprà, 2010) and forecasting research is positioned with forecasting being an end in itself (e.g., Makridakis et al., 1998). At level 1 lies literature that recognises the existence of the other field but does not engage with it. For example, inventory research will note the need to forecast demand parameters (but do not), see, e.g., Waters (2008); forecasting research will touch on inventory implications but will not explore forecasts' utility, see, e.g., Ord and Fildes (2012) and Ord et al. (2017). This body of literature departs from level 0 by recognising either the need to forecast demand parameters, or the fact that the forecasts are ultimately going to be used for inventory control.

Level 2 describes literature that takes the first steps towards integration. For forecasting, integration begins when forecasts are judged on bottom line inventory considerations and metrics (i.e., considered a means to an end). For inventory control, it begins when one uses forecasts of demand (either, e.g., type of distribution and its moments) as opposed to assuming demand is known. At level 2, we consider the 'serial' application of forecasting and inventory control (e.g., Fildes and Beard, 1992; Willemain et al., 1994; Eaves and Kingsman, 2004; Syntetos and Boylan, 2006). Regardless of where the particular focus lies, they tend to estimate demand parameters and then employ inventory control policies, recognising the need for integration.

At level 3 we consider integrated deliberation of inventory forecasting problems. Here, demand characteristics are forecasted and used to in-parallel understand and/or educate the inventory forecasting model or our understanding of it. Authors grapple underlying attributes affecting the system, such as correlation in demand (e.g., Lagodimos et al., 1995; Graves, 1999) or forecasts (e.g., Johnston and Harrison, 1986; Prak et al., 2017) and/or evaluate their model based on an overarching metric of inventory performance. That is to say, any selection or decisions on both the forecasting method and the inventory policy are based on the total performance of the system rather than the performance of its constituents. (This does not subtract from the importance of measuring forecast accuracy, as it will always be relevant for tracking and monitoring the performance of a system.)

Beyond the integration levels (and to also help us decide on it), further information has been extracted from each paper: The forecasting methods and inventory policies employed, forecasting and inventory performance metrics, methodological and contextual information. The process of individual paper evaluation is graphically illustrated in a flowchart (Figure 3, this being also the flowchart 'explosion' of the process "paper classification" in the review protocol depicted in Figure 5 of the next Section 2.2):

2.2. Keyword selection and search

Our keyword sets were informed in three ways: i) initial keyword selection and search, ii) (key)word analysis on the result-

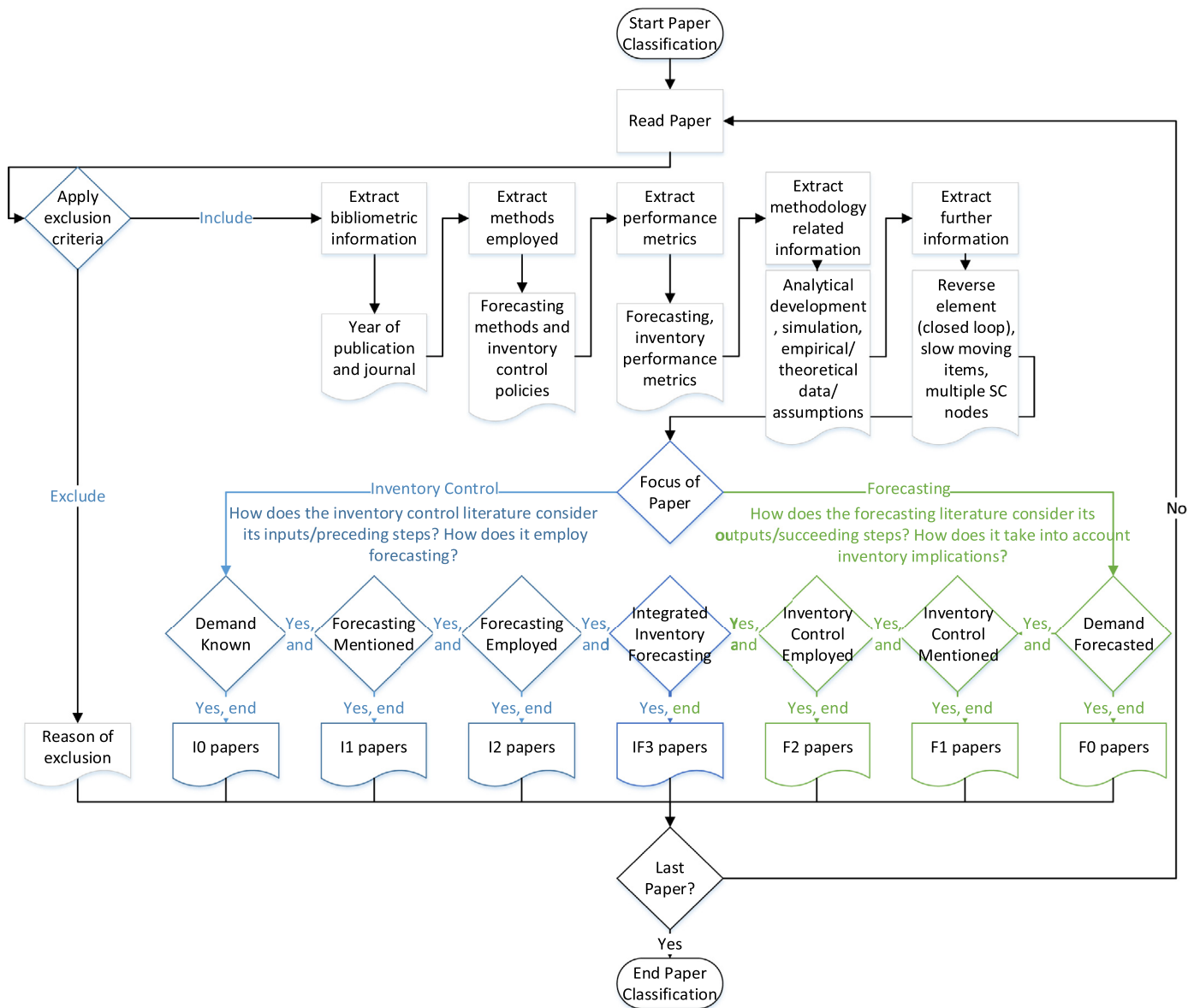


Fig. 3. Paper classification flow chart.

ing paper (pre-)sample, and iii) expert consultation. Details of this process can be found in Appendix A. The final keywords used are shown in Figure 4.

This keyword string provides our initial sample of 880 papers, as returned by Scopus. Scopus employs automatic indexing and does not provide a way to ignore the computer-generated keywords it associates with each paper⁶. There is a risk associated with automatic indexing, in potential automatic misclassification of articles (also see Section 2.3). After manually excluding papers that ended up in the initial sample by merit of the automatic indexing alone, 502 papers remained, which were all read and classified (or manually excluded; see Figure 5 for the review protocol).

2.3. A note on keyword string database searches

Before we move on to discuss the integration framework, some important notes are warranted with regards to keyword-based

database searches. Considerable effort has been expended to construct a search string of keywords, to focus the sample on the matter investigated, while trying to not omit relevant parts of the literature. However, no string is perfect, nor everything always works as intended with database searches. Both of these factors have consequences in the constancy of the returned sample.

Irrelevant (erroneous) article inclusions need to be kept to a minimum, to reduce the size of the sample and manual intervention (that will be expended to manually exclude them later). Relevant article exclusions (false exclusions) should be minimized as well, to not miss parts of the literature. These are often competing goals when compiling the search string, an iterative process of trial and error. Attempting to exclude all irrelevant articles will result in missing parts of the relevant literature. Conversely, attempting to not miss a single relevant paper will result in too big a returning sample, with many irrelevant articles. The goal of minimising both erroneous inclusions and exclusions is at times competing, and a delicate balance needs to be reached through compromise.

By merit of looking for papers that concurrently deal with forecasting and inventory control, works that discuss themselves in

⁶ We need to differentiate here between keywords the authors have elected to represent their work ('author keywords'), and keywords attributed to each paper by, in this case, SCOPUS algorithms ('index keywords').

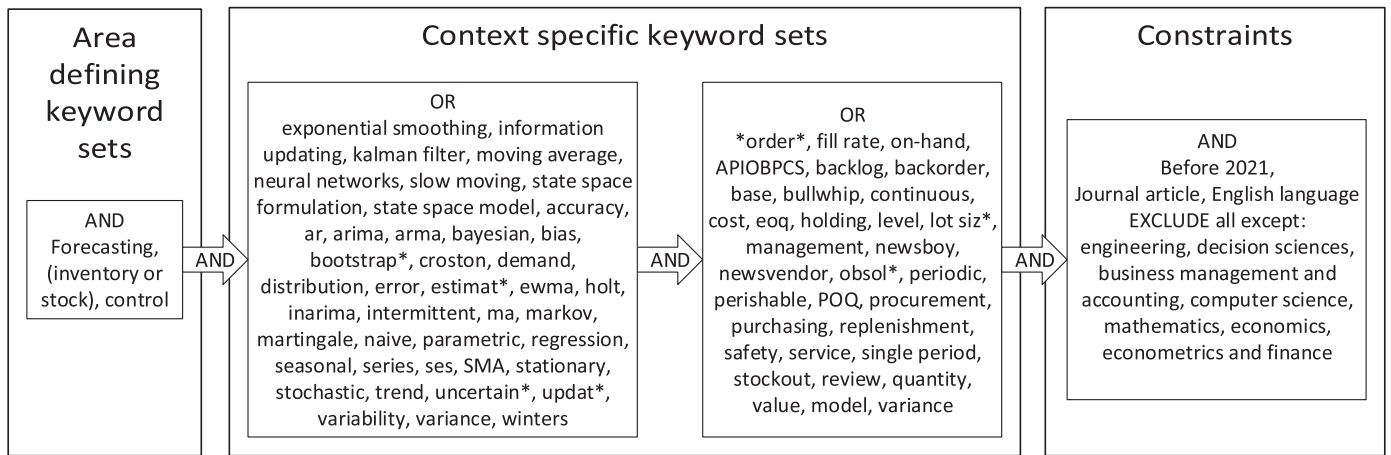


Fig. 4. The search string. We review articles that have forecasting, inventory or stock control, and at least one from each of the context-specific keyword sets, in their titles, abstracts or author-selected keywords.

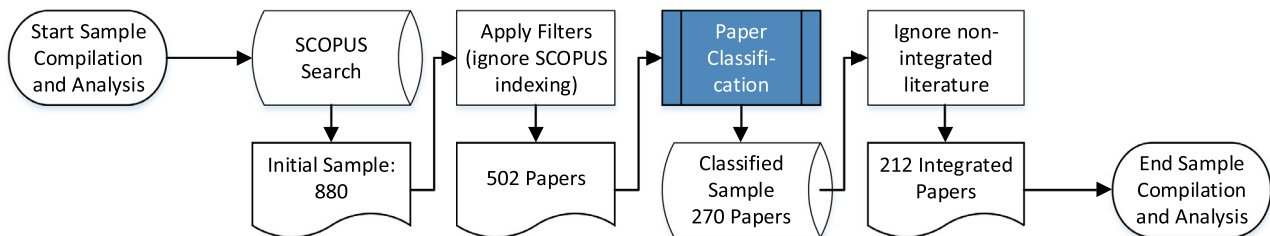


Fig. 5. Review protocol. The paper classification process (blue) was discussed in Figure 3. Scopus returned 880 papers. After ignoring the database’s automatic indexing, 502 articles were read and 270 were found relevant. Of the 270 articles, 212 were of levels 2 and 3. Most of the following analysis is conducted on this final sample of integrated literature.

terms other than forecasting and inventory/stock control might be erroneous exclusions. This includes top-level, methodological (often older) approaches from fields such as statistics. We attempt to right this in Section 4, where we partly depart from the confines of the sample to also include other relevant papers that might have been missed. It is important to note here that we do not attempt to compile (or claim to present) an exhaustive list of all the integrated articles ever written. We attempt to capture an objective, reproducible snapshot of the literature (so it could be repeated in the future and see whether things have changed), and explore which streams have shown promising venues toward integration.

3. Sample analysis

Following the doctrine described in Section 2, we read every article and decided on whether it should be included, its integration level and further information. Here we present the results of this process. We close the section with an attempt at synthesis of the results, which we use as a springboard for further discussion in the subsequent sections.

3.1. Integration levels classification

Out of 270 articles reviewed, five articles were found to be level 0 and 53 articles were found to be level 1. Since we are interested in the integrated literature, these articles are removed from the final sample as wrongful inclusions. The subsequent analysis is constrained to the integrated literature (212 articles) of levels 2 (118 articles) and 3 (94 articles) (see Figure 6).

3.1.1. Integration over time

The first papers in this area started appearing in the beginning of the 1970s. In Figure 7, we plot all integrated articles against

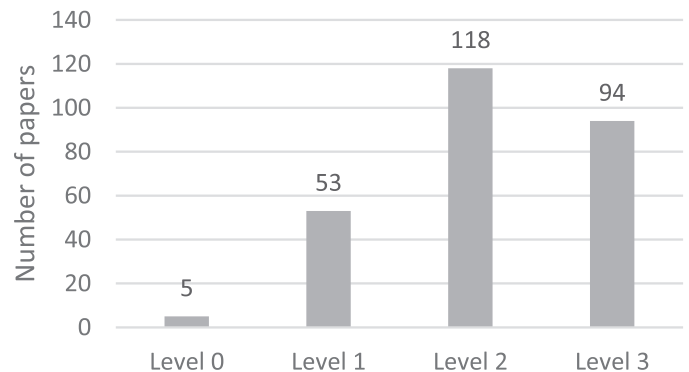


Fig. 6. Integration level breakdown of the sample of 270 articles. Levels 0 and 1 are highly underrepresented due to integration-inclined keyword search. We focus on the 212 articles of integrated research (articles of integration levels 2 and 3).

time, biannually. We can see there has been a sustained increase in the total number of papers published annually since then. We note a step-change in the late 1980s and early 2000s. The area really started growing in the late 1980s but has plateaued at about 10 articles per year since its peak at 2010 (which was mainly driven by a special issue in the area – see next section). One interpretation could be that since that special issue, some authors preferred to focus on one of the two sides.

In Figure 8, we overlay areas of the relative growth for each of levels 2 and 3 (left) and their percentage split (right), between integration level 2 and 3 papers over the last 20 years (again biannually). We can see that the “mixture” of papers is changing towards more integrated approaches over time. Altogether, we see a slight relative increase in the number of level 3 papers.

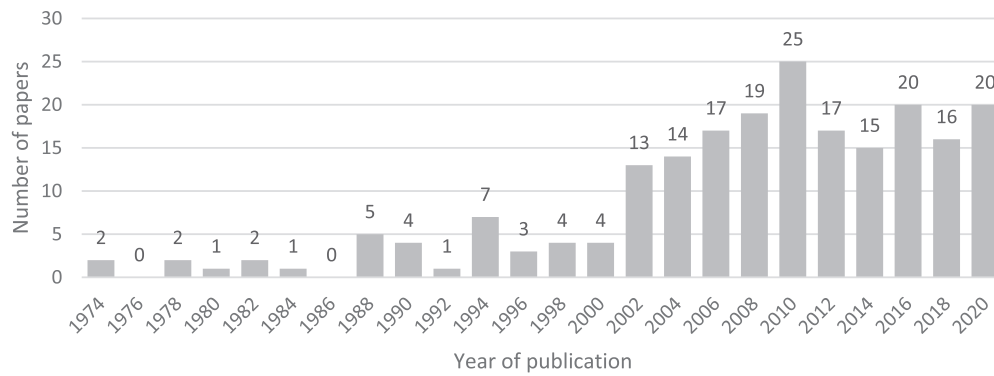


Fig. 7. Publications on integrated literature (levels 2 and 3), biannually. We see a total increase in the literature, reaching almost 10 articles per year in the last fifteen years.

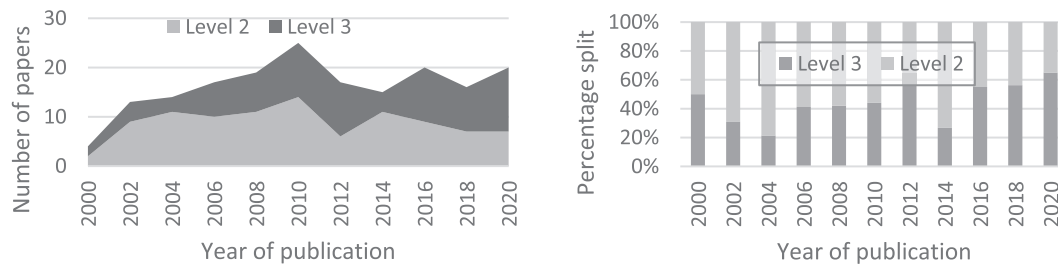


Fig. 8. Publications per year on integration levels 2 and 3 (stacked graph of absolute numbers, left; relative split, right). There is a slight relative percentage increase of papers of level 3 since 1999.

Integration level 2 identifies the focus of the paper lying either in forecasting (F2, 49) or inventory control (I2, 69). This slight prevalence of focus in inventory hints that most of the integrated literature focus lies in inventory.

3.1.2. Journal titles

Where is this research published? In Figure 9, we plot the level of integration against the journal of publication. Understandably, different journals publish different numbers of issues each year, and therefore there are journals that publish many more articles than others. This of course affects these results. For example, the International Journal of Production Economics (IJPE) published 230 articles per year, almost four times more than the International Journal of Forecasting (IJF) at 63 articles per year, but almost half than the European Journal of Operational Research (EJOR) at 414 articles per year, on average⁷. Taking this into account and with some exceptions, integrated literature is quite spread out.

In Figure 10, we focus on the three most represented journals of our sample. In IJPE, we can see that after the first articles appeared in 1992–1995, there is an increase in the number of published inventory forecasting literature - in particular, after 2008 (introduction of the inventory forecasting stream in ISIR). A peak appears in 2010 with the publication of a special issue entitled “Supply Chain Forecasting Systems”, which seems to have supported a further growth since. In contrast, integrated research in the IJF seems constant over time, while EJOR exhibits a modest growth which is perhaps corrected downwards over the last 10 years.

3.2. Forecasting

3.2.1. Forecasting methods

In Figure 11, left, we can see the forecasting methods and approaches of the integration literature, across levels 2 and 3. Ex-

ponential smoothing is the most popular approach to forecasting in our sample, with simple moving averages, Croston-like methods (also exponential smoothing-based) and ARIMA following. The use of exponential smoothing and simple moving average methods’ (see Hyndman and Athanasopoulos, 2018) wide adoption can be partially explained by the fact that they are often used as benchmarks (sometimes concurrently, e.g., Eaves and Kingsman, 2004). On the right, we see the relative frequency for a method to be applied in level 2 and level 3 literature where a method is more probable to appear.

Entire distribution-based methods (e.g., Kolassa, 2016) seem to be well suited for integrated approaches. The same applies for direct quantile estimation. Taylor (2007) proposes an exponentially weighted quantile regression as a treatment to highly volatile and skewed daily time series. Amrani and Khmelnitsky (2017) estimate quantiles by attributing weights to samples based on their chronological order for non-stationary demand patterns. Cao and Shen (2019) find improvements in quantile estimation, when compared to Holt-Winter’s forecasts coupled with a normality assumption of residuals, in two seemingly well-behaved empirical time series.

When distributional assumptions are reasonable, Bayesian methods can offer a way to incorporate unknown demand parameters in an inventory decision model (Prak and Teunter, 2019), as new information becomes available. They can be perceived as the middle ground between assumed known (and unchanged) demand distributions on the one side and data driven distribution-free approaches on the other, producing full predictive distributions. Yelland (2009) use Bayesian state-space formulations to forecast low-count (intermittent) demand. Wang and Mersereau (2017) formulate a Bayesian inventory problem after change-points in demand and provide heuristics to estimate its parameters.

A number of heuristics have been employed in this literature to tackle complex integrated issues, when problems become analytically intractable. We opt to present them under forecasting, as they are used to forecast quantities such as reorder point/order-up-to levels (e.g., Power approximation, Naddor’s Heuristic, Normal ap-

⁷ Per year averages across time of circulation per journal. Totals were taken from searches in Scopus and are pertinent to their respective coverage years. IJPE: 6,873 since 1991; IJF: 2,211 since 1985; EJOR: 17,802 since 1977. Search up to 2020.

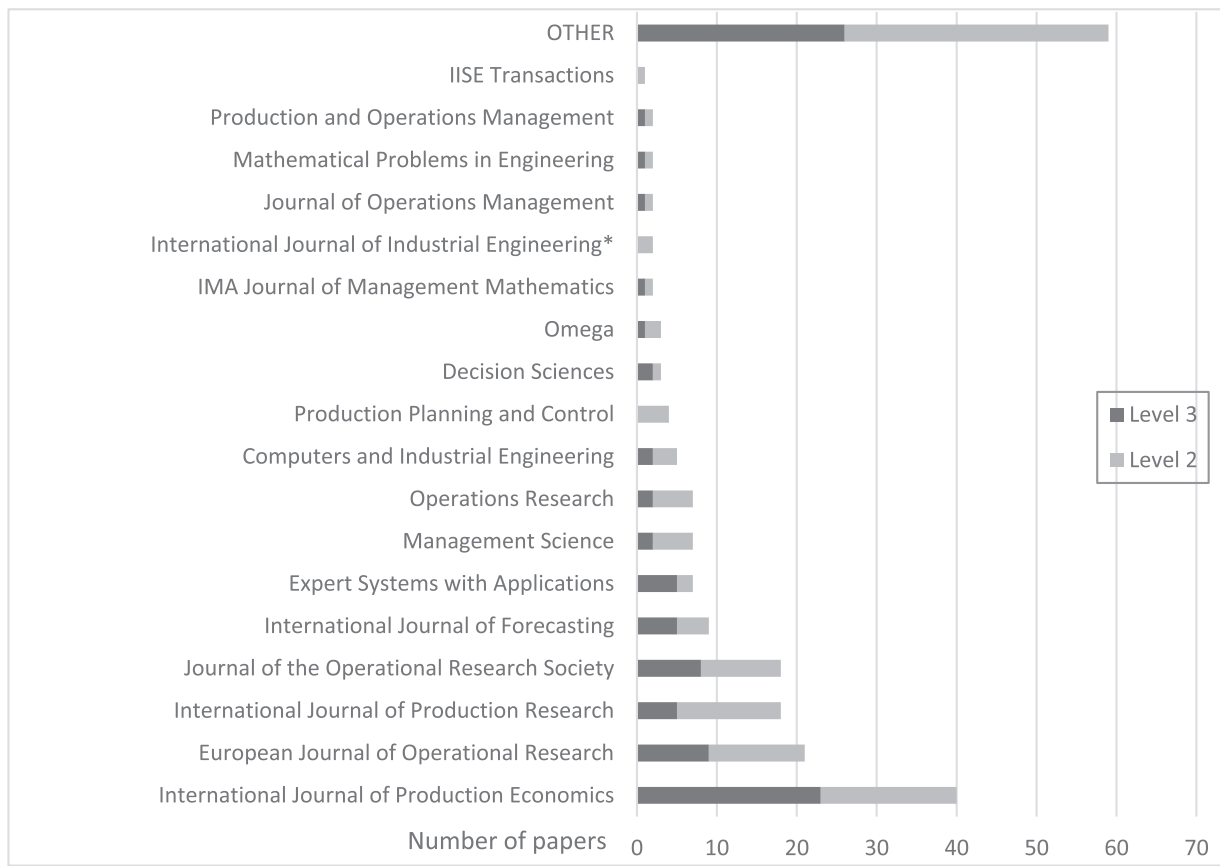


Fig. 9. The dispersion of the sampled integrated literature across journals. *full journal title: "International Journal of Industrial Engineering: Theory Applications and Practice".

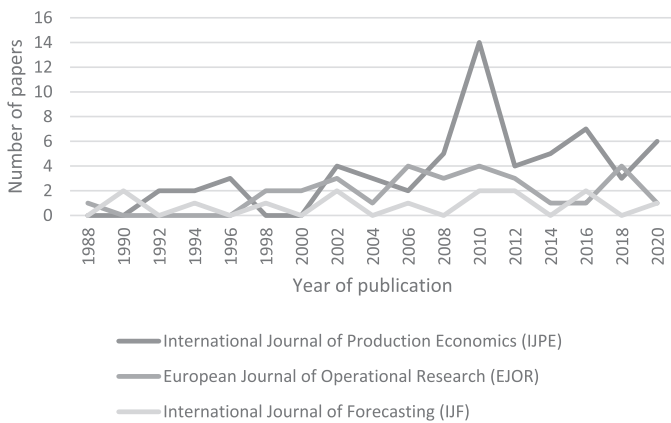


Fig. 10. Integrated (levels 2 and 3) inventory forecasting articles in IJPE, EJOR and IJF. In 2010, IJPE published a special issue on "Supply Chain Forecasting Systems".

proximation in, e.g., Sani and Kingsman, 1997; Babai et al., 2010). They include evolutionary algorithms such as the ant colony optimisation (e.g., Su and Wong, 2008) the hybrid artificial bee colony-chaos algorithm (e.g., Tang et al., 2020), or approximate Bayesian estimator smoothing heuristics (e.g., Karmarkar, 1994), among others.

Machine learning (ML) approaches have gained popularity in recent years (12 out of 19 such publications in our sample are from the last five years) in bypassing finding the distribution of demand to integrate parameter estimation and inventory optimization. One of the things ML allows is the incorporation of 'features' or covariate information about the products. Oroojlooyjadid et al. (2020) ap-

ply data-driven deep learning to a newsvendor formulation with multiple features. Ban et al. (2019) propose another data-driven approach with covariate information (e.g., price, colour, etc.), the 'residual tree method' (an extension of the scenario tree method), a combined forecasting and optimization algorithm to choose order quantities. Cao and Shen (2019) employ a 'double parallel feed-forward network-based quantile forecasting' neural network to directly estimate quantiles for a newsvendor formulation, also for new items.

The 'other' category includes forecasts based on diffusion models (e.g., Ho et al., 2002), Markov chain formulations (e.g., Cervellera and Macciò, 2011), failure rate calculations (e.g., Ghodrati and Kumar, 2005), judgement (e.g., Syntetos et al., 2009), and robust optimisation approaches (e.g., Kim and Chung, 2017), among others. Please note that these are not strictly forecasting methods, but rather describe procedures and formulations that are used either to produce forecasts or to estimate inventory parameters. Of these, robust optimisation seems to be very close to the subject of integration. 'Robust' refers to robustness against distributional assumptions, meaning that the methodology provides distribution-agnostic interval forecasts.

3.2.2. Forecasting performance metrics

When it comes to measuring the performance of the forecasts, we can see that (mostly mean) squared errors and (mostly mean) absolute errors are the two most popular error metrics (Figure 12). Both of these metrics have links with inventory control in terms of the calculation of safety stocks. It is noteworthy that the vast majority of metrics employed emphasizes point forecasts of the mean demand, with squared errors as proxies for the demand variance (that would in most cases define safety stocks).

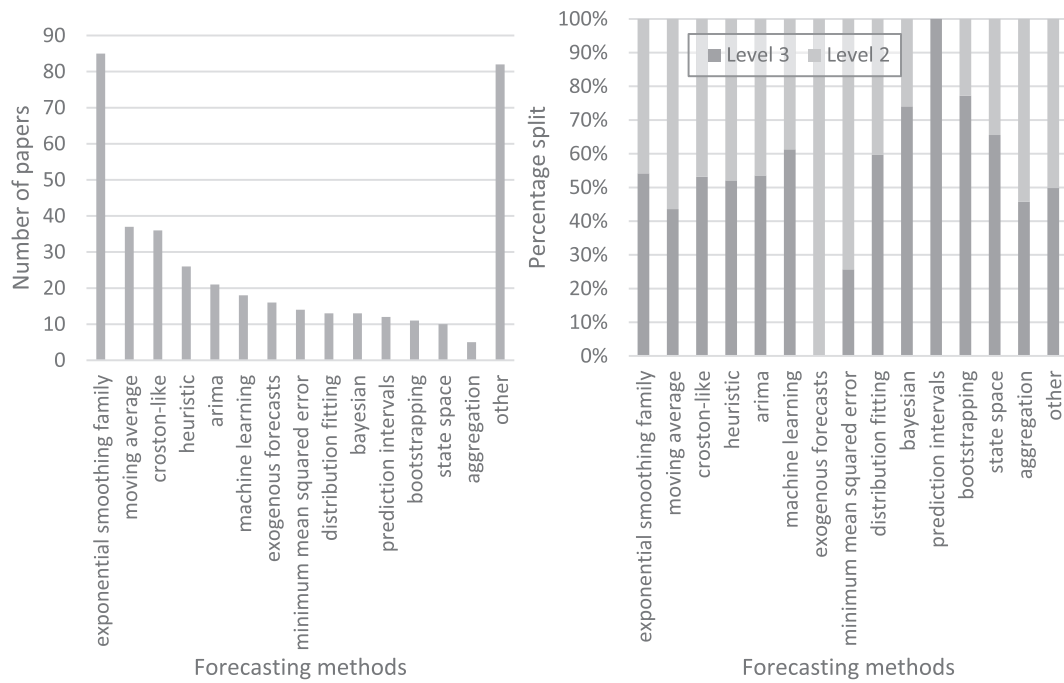


Fig. 11. On the left, forecasting methods employed by the literature for integration levels 2 and 3. Prevalence of exponential smoothing, moving average and Croston-like methods. On the right, (normalised, relative) preference of method across the integration levels. Distribution-based methods (bootstrapping, prediction intervals), Bayesian and state-space formulations closer to level 3. Multiple entries allowed per paper.

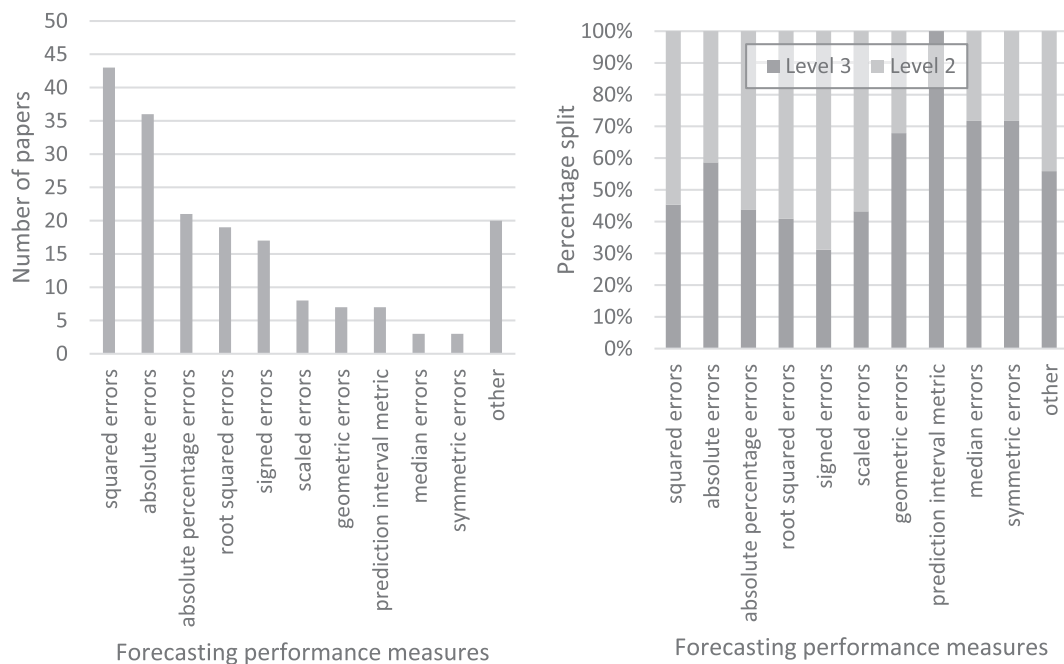


Fig. 12. On the left, forecasting performance measures employed by the literature at levels 2 and 3. Squared and absolute errors are most prominent. On the right, relative preference of errors across the integration levels. Multiple entries allowed per paper.

While there are developments in point forecast accuracy estimators (e.g., [Petropoulos and Kourentzes, 2015](#)), there is a need to develop methods that can judge the accuracy of the entire lead time demand distribution or at least percentiles of interest ([Kolassa, 2016](#)). Such methods include the (discrete or continuous) ranked probability score and probability integral transforms ([Yelland, 2009](#); [Kolassa, 2016](#)). The category ‘other’ includes relative errors (e.g., [Willemain et al., 1994](#)), tracking signals (e.g., [Tiacchi and Saetta, 2009](#)), and times best (e.g., [Chatfield and Hayya, 2007](#)), among other forecasting performance measures.

3.3. Inventory control

3.3.1. Inventory policy

When it comes to inventory policies, there is a clear dominance of the simple yet robust periodic order-up-to (T,S) policy (113 articles), with the continuous re-order point, order quantity (r,Q) policy following (29 articles) (see [Figure 13](#)). The (T,S) policy in this literature, refers to fixed periods (T), and therefore T does not need to be optimised. When one is only seeking to optimise the order-up-to level S, the problem degenerates to a solution of a simple

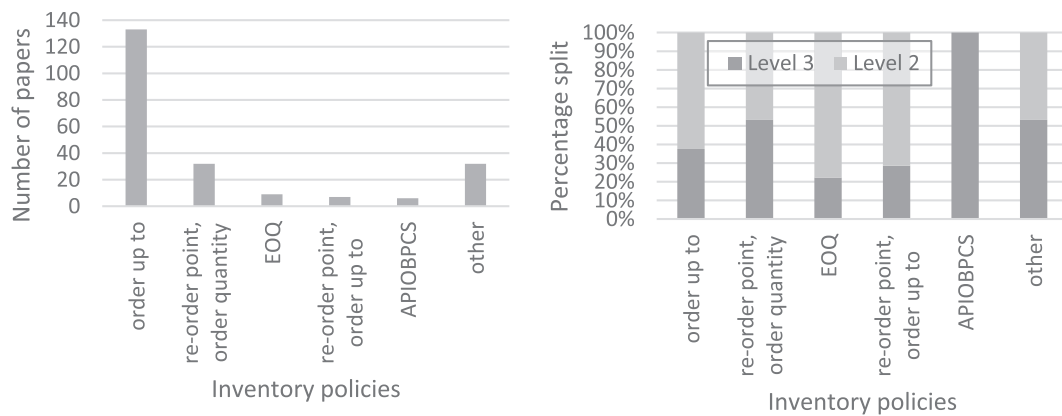


Fig. 13. On the left, inventory policies measures employed by the literature at levels 2 and 3. On the right, relative preference of policies across the integration levels. Multiple entries allowed per paper.

news vendor problem which essentially provides a target service level to be achieved (Eppen and Schrage, 1981). The large number of papers working with the (T,S) policy is easily explained by taking into account the simplicity of the model, both to analyse and simulate, offering to mitigate some of the inherent complexity of the integrated approaches. It is also interesting to note that when optimised in all variables, the (T,S) policy performs near optimally compared with all other periodic review policies (see Lagodimos et al., 2012, for an in-depth discussion and de Kok, 2018, who recently reached the same conclusion with a different approach). On the other hand, the (r, Q) policy is quite more convoluted (see Zheng, 1992, for its basic analysis under simple stochastic demand assumptions).

The prevalence of periodic order-up-to policies in the integrated field may be partially attributed to the fact that a) in reality almost every inventory system is periodic, as forecasts are also almost always periodic in nature (be it in hours, days, weeks, months), b) the move from the continuous to the periodic domain is a move towards more realistic (integrated) demand representations (e.g., Lagodimos et al., 2018), and c) the robustness of the order-up-to policy.

We note the contributions from the (almost by definition integrated) control theoretic approach Automatic Pipeline Inventory and Order Based Production Control System APIOBPCS (John et al., 1994, see Lin et al., 2017 for a recent review of its applications). APIOBPCS is a feedback-based block diagram framework consisting of forecasting, work in progress, inventory, and production lead time policies (controllers). Appropriate parameters are selected among the policies with the “competing objectives of 1) rapid inventory recovery and 2) attenuation of the unknown demand fluctuation [...] in an effort to understand supply chain dynamics more completely” (Lin et al., 2017, pp. 136–137). Clearly, this is in line with the goals of integrated inventory forecasting.

The category ‘other’ includes fuzzy methodologies (e.g., Gumus et al., 2010), empirical/rule-based approaches (e.g., Lee and Liang, 2018), and direct use of (forecasted) quantiles to estimate inventory parameters of interest (e.g., Amrani and Khmel'nitsky, 2017), among others.

3.3.2. Inventory performance metrics

Most works in our sample (see Figure 14) measure inventory performance through a (minimized) cost function (including, in most cases, inventory holding and backorder costs) or customer service level achieved. (Maximized) profit functions (e.g., Johnston et al., 2011) and fill rate service level representations (e.g., Heath and Jackson, 1994) are less popular alternatives. Cycle (or customer) service level (service level α) is very easy to find and com-

pute since it corresponds to a simple probability expression, while the fill rate corresponds to a more complicated form (see Silver et al., 2017, Schneider, 1981, for an in-depth exposition of this service level measures; and Diks et al., 1996 in a multi-echelon context). Some articles avoid the cost/profit representation by measuring directly average inventory, backorder volumes, or stockouts in terms of units (e.g., Babai et al., 2014).

Various inventory-related variance metrics (e.g., bullwhip effect, Dejonckheere et al., 2002, or net stock amplification, e.g., Jaipuria and Mahapatra, 2014), closely associated with, but not constrained to, system dynamics or control theoretic approaches (e.g., APIOBPCS) can be considered inductive to integration, in as much as they benchmark inventory performance against demand- or forecast-based characteristics.

The category ‘other’ includes transportation costs (e.g., Tiacchi and Saetta, 2009), expected waiting time (e.g., Zhu et al., 2017), overtime and subcontracting (e.g., Ha et al., 2018), and expediting (e.g., Clottey et al., 2012), among others.

3.4. Further information

There are 123 papers that include theory development of some/any kind, while 139 papers employ simulation, see Figure 15, left. Simulation has played an integral role in revealing inefficiencies of the traditional assumptions and approaches when applied in combination with real data (see Cattani et al., 2011), and in providing arguments for integration (e.g., Eppen and Martin, 1988). At level 3, the split between these two categories is almost equal, whereas at level 2, we find a higher proportion of papers employing simulation (58%). Out of the 139 papers that employ simulation, there are 90 papers that do so using empirical data (e.g., Eaves and Kingsman, 2004, that use aircraft spare part monthly time series; or parameters taken from the real world), and 54 using theoretical data (hypothesised, e.g., drawing from a normal distribution, e.g., Zhao and Leung, 2002). In total, 80 papers employed empirical data and 100 papers employed theoretical data (Figure 15, middle).

There is another vector of integration that takes the perspective of the entire supply chain, trying to investigate or optimise key performance indicators across it (Figure 15, right, “multiple nodes”). 43 (65% of which at level 2) articles in our sample did so, concerning themselves with more than one member (or node) of a supply chain (see de Kok et al., 2018 for a recent review of the area). Intermittent demands refer (as mentioned) to patterns where periods of positive demands are scarce, interspersed around successive periods of no demand. Our sample contains 63 papers dealing with such “slow” moving items, almost evenly split be-

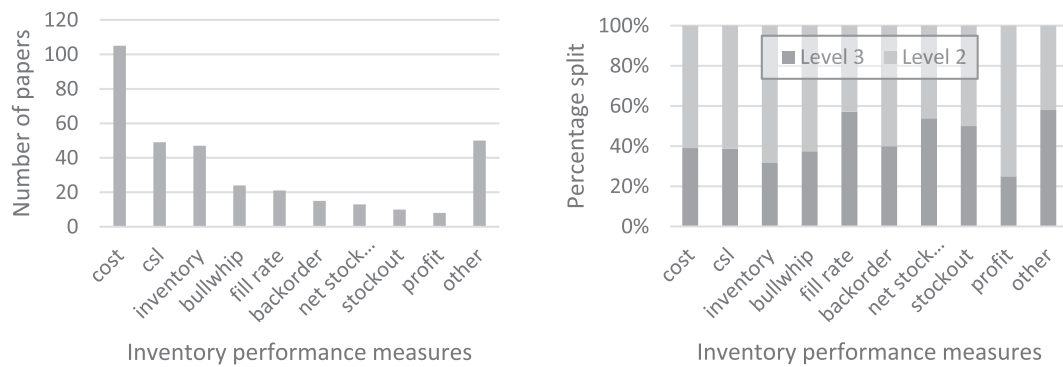


Fig. 14. Inventory performance measures employed by the literature at levels 2 and 3. On the right, relative preference of policies across the integration levels. Multiple entries allowed per paper.

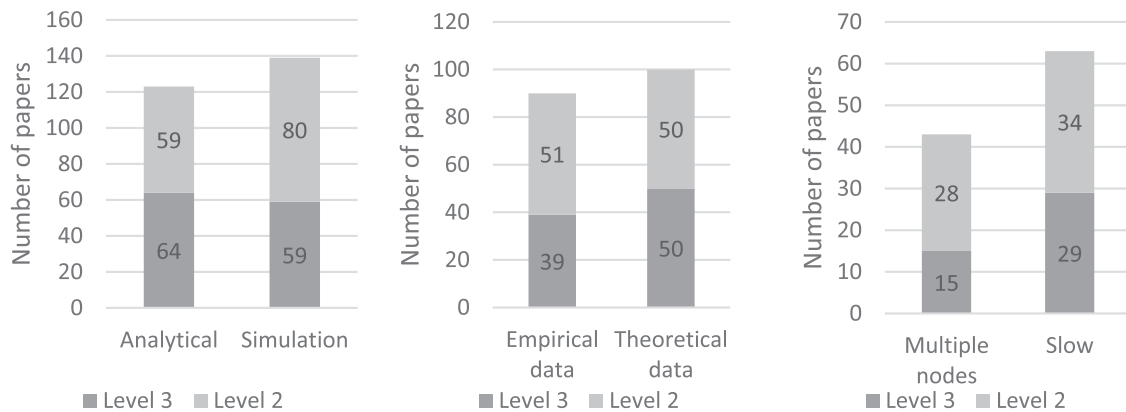


Fig. 15. Number of papers that include analytical developments, simulation, empirical and theoretical data, multiple nodes and slow moving items. Integration levels 2 and 3.

tween levels 2 and 3 (Figure 15, right, “Slow”). This high (proportionally to the sample size) number of papers highlight the role intermittent demand research has played in exercising and promoting inventory forecasting integration (see Boylan and Syntetos, 2021).

Finally, three articles considered closed loop supply chains (see Goltso et al. 2019b for a recent review of the area). While the sample has missed a few others (e.g., Toktay et al., 2000), closed loop inventory forecasting is an interesting niche open to contributions. It has been noted that while returns forecasting more often than not considers forecasting (returns) and inventory control jointly, as the focus lies on returns, literature in the area has almost invariably assumed known demand (Goltso et al., 2019b).

3.5. Summary and synthesis

We find that the most common combination of forecasting procedures with inventory policies is that of exponential smoothing and moving average with order-up-to policies. The mean forecasts are then coupled with variance estimations and assumed distributions of errors to compute percentiles of interest. The simplicity of these constituent formulations facilitates the discussion of the inherently more complex integrated inventory forecasting question. Often, but perhaps not often enough, these combinations form the basis of rigorous benchmarking of more complex proposed approaches.

We note that simulation has played a very important role in the development of arguments in the integrated literature (as it has in supply chain management in general, see, e.g., Fagundes et al., 2020). Firstly, it has exposed the mismatch of various theoretical assumptions with the reality faced by practitioners, especially

when coupled with empirical data sets coming from industry. Secondly, when properly employed, simulation can show the practical impact of proposed methodologies when compared to simpler benchmarks.

A number of promising approaches and areas emerge from our sample: quantile estimation, robust optimisation, bootstrapping, Bayesian inventory control, data driven and machine learning and importantly the performance measurement of quantile and density forecasts. We expand on these streams in the following section.

4. Discussion

Based on the review of our sample in Section 3, we select a few promising areas for integration and attempt a more detailed discussion. Some of the papers presented below are from the sample, but, in general, papers presented here are not constrained to it (especially when it comes to historic methodological developments that for reasons discussed in section 2 do not appear in our sample). Some come from forward and backward ‘snowball’ searches, complemented by independent mini reviews of the relevant streams.

4.1. Simple but fundamental interventions

Before getting into such streams however, it is important to note that integration and improvements can also originate from simple (yet fundamental) interventions. The seminal intervention to forecasting for intermittent demand came from Croston’s (1972) appreciation of the ‘issue point bias’, and his proposed solution to it: forecast (smooth) demand size and length of inter-demand intervals independently (and update only after a demand

occurrence). Syntetos and Boylan (2005) quantified and approximately corrected a positive bias in Croston's method. Further elaboration in the area and the problem of obsolescence in particular has prompted further important interventions (e.g., updating the demand occurrence every period in Teunter et al., 2011; see also different approaches by Prestwich et al., 2014 and Babai et al., 2019).

Another example of integrated thinking providing important interventions in the area comes from the realisation that demand parameters cannot be directly substituted by forecasted ones (as is common). Prak et al. (2017) show that for mean-stationary normally distributed demand, forecast errors are auto-correlated and thus safety stocks need to be inflated to avoid understocking. Prak and Teunter (2019) provide a Bayesian framework to incorporate this parameter estimation uncertainty, one that can be applied to any inventory model, demand distribution and parameter estimator. Prak et al. (2021) provide a closed form solution to calculate compound Poisson demand parameters, robust to the compounding distribution misspecification. They find improvements in the finite-sample bias and achieved fill rates in a continuous review order-up-to system against literature suggested method-of-moments and maximum likelihood formulations.

These select examples show there is certainly tremendous scope in uniting inventory and forecasting by a better conceptual understanding of their interaction, and identification of relevant opportunities for improvement. While more recent data driven and machine learning approaches, covered in the following subsections and elsewhere, increasingly find applications to inventory forecasting problems, we have not yet exhausted the simple interventions that offer direct solutions but also help speed up computations. A further important benefit brought through such (often closed form) solutions to fundamental issues of inventory forecasting is their ease of communication, brought by their transparency (white-box nature – which also shortens potential innovation-adoption gaps).

4.2. Quantile estimation

The general aim of integrated inventory forecasting is to derive optimal inventory parameters without resorting to dubious distributional assumptions (an idea tracing back to at least Iyer and Schrage, 1992, for the deterministic (s, S) system). A number of papers have bypassed the problematic normality (or other distributional) assumptions by directly forecasting the quantile of interest. This stream of literature started with (was inspired by) Koehler and Bassett's (1978) work on (linear) quantile regression (extended to the autoregressive case by Koehler and Xiao, 2006).

Trapero et al. (2019a) employ kernel density estimation (KDE – Silverman, 1986) and (generalised) auto regressive conditional heteroskedasticity (ARCH – Engle, 1982; GARCH – Bollerslev, 1986) to forecast quantiles for directly computing safety stock levels. They find improvements in terms of inventory performance, via KDE in shorter lead times when the normality assumption is most suspect, and via GARCH in longer lead times where conditional heteroskedasticity becomes dominant. In a subsequent work, they find further improvement when combining such quantile forecasts (Trapero et al., 2019b). Taylor (2007) and Cao and Shen (2019) find significant accuracy improvements in estimating quantiles of interest, when compared to more traditional approaches (normality assumption centred on simple exponential smoothing and Holt-Winter's point forecasts, respectively), which they traced back to the unsuitability of the traditional distributional assumptions.

Of course, another approach is to forecast the entire distribution and then to extract desired quantiles of interest (Fildes et al., 2019; see, e.g., Gneiting 2011b; Kolassa, 2016; Sillanpää and Liesjö, 2018). This move from point to probabilistic forecasts is cutting across a

number of scientific disciplines (see Gneiting and Katzfuss, 2014, for a review and further argumentation).

4.3. Robust optimisation

Robust optimisation (RO) deals with uncertain variables by only looking at intervals without a need for further distributional information (Wei et al., 2011). In this sense, it produces results regardless of the true underlying distribution that generates the data. RO was proposed by Soyster (1973), and it has been criticised, and for years dismissed by researchers, because the robustness originates from assuming worst-case-scenarios for all parameters, (see arguments in Ben-Tal and Nemirovski, 2000). To alleviate the overly conservative nature of the results, Mulvey et al. (1995) introduced scenario-based RO, while subsequent research applied RO to linear programming problems with uncertainty sets (Ben-Tal and Nemirovski, 1998, 1999, 2000 and independently by El-Ghaoui and Lebret, 1997; El-Ghaoui et al., 1998; see Gabrel et al., 2014; Yanikoglu et al., 2019, for reviews of the area). RO comes with the benefit of computational tractability over alternative methods (Ben-Tal et al., 2009).

This independence from demand distribution assumptions has made the RO approach conducive to inventory control applications under demand uncertainty. Bertsimas and Thiele (2006) apply RO to an (s, S) inventory setting with backorders, and find evidence that it outperforms dynamic programming formulations. See and Sim (2010) propose a RO approach to address a (T, S) setting facing ARIMA (0,1,1) demand, and find that their formulation performs 'reasonably well' compared to optimal policies, despite using significantly less information. Thorsen and Yao (2017) were the first to also consider uncertain lead times in this setting. Bertsimas et al. (2019) provide an adaptive RO framework to address dynamic problems, providing the ability for it to adapt as new information becomes available. The above works all find improvements over mis-specified optimum policies (the cases where real or realised distributions are different than the assumed or sampled ones).

4.4. Bootstrapping

Bootstrapping is a data driven non-parametric method of constructing empirical distributions of demand, and it works by resampling from the historical demands (Efron, 1979). It is a generalisation of the jackknife method (Quenouille, 1949, 1956; see Miller, 1974 for a review). While we cover more data driven methods in Subsection 4.6, we dedicate a subsection to bootstrapping for its importance in inventory forecasting. Applications of bootstrapping include Clements and Taylor (2001) for autoregressive models, Snyder et al. (2002) for exponential smoothing models, Rubin (1981) for Bayesian models (sampling from the posterior distribution rather than the observed data). Bertsimas and Sturt (2020) consider deterministic algorithms to calculate exact bootstrap quantities for the sample mean and confidence intervals.

The first application of bootstrapping to inventory control can be traced back to Bookbinder and Lordahl (1989), who used it to compute the reorder level in an (s, Q) system against a cycle service level criterion. Fricker and Goodhart (2000) apply bootstrapping to the calculation of re-order points for a (s, S) system against a fill rate and other service criteria. Bootstrapping has found relatively wide application to intermittent demand contexts, where the scarcity of positive demand occurrences make parametric approaches harder to implement (e.g., Willemain et al., 2004; Viswanathan and Zhou, 2008; van Wingerden et al., 2014; Hasni et al., 2019b; see Hasni et al., 2019a for a recent review on bootstrapping in intermittent demand contexts).

While bootstrapping offers a solution to lacking distributional fits, it does not solve the issue of uncertainties arising from small

samples, nor of relevant overfitting issues. If only few observations are available, then bootstrapping draws from only these observations, and does not anticipate on possible future values that are different. Having said this, exceptions such as the jittering process from [Willemain et al. \(2004\)](#) allow expectations of demand sizes not previously observed.

4.5. Bayesian analysis

Bayesian analysis provides a formal way to incorporate prior information, at times even before data becomes available to the demand forecasting process. The forecaster can use domain knowledge to select a prior distribution, which is updated through the Bayes theorem into the posterior distribution as more data become available. Applications of Bayesian theory to inventory control were pioneered by [Dvoretzky et al. \(1952\)](#), [Scarf \(1959\)](#), and [Azoury \(1985\)](#).

This approach could be seen as a sound way to utilise classic inventory theory using a nimbler approach that updates the parameters of the distribution as more data becomes available. The dependence on distributional assumptions also makes the Bayesian approach resilient to issues of overfitting that can be generally found in bootstrapping and machine learning approaches. A problem with this approach is that realistic distributions with more than one parameter often lead to intractable solutions, and are often treated with unrealistic assumptions (e.g., Normal distribution with known variance, in [Azoury and Miyaoka, 2009](#) and [Chen, 2010](#)). This assumption is problematic in inventory control as the variance dictates the size of the safety stocks and therefore is quite (very) important ([Prak and Teunter, 2019](#)).

Bayesian theory is very useful in machine learning approaches, and widely used in a number of neural network (and other) applications. For example, [Boutselis and McNaught \(2019\)](#) employ a Bayesian Network machine learning approach to a service logistics context.

[Prak and Teunter \(2019\)](#) use Bayesian theory to discuss how demand uncertainty should be taken into account in inventory control (when inventory formulations use estimated – rather than known – parameters of demand). [Toktay et al. \(2000\)](#) and [Clotey et al. \(2012\)](#) use Bayesian updating to incorporate new information on the returns of used products in distributed lag model formulations, as such products returned. In an intermittent demand context, [Babai et al. \(2021a\)](#) propose a compound-Poisson approach, while [Ruiz et al. \(2021\)](#) employ Bayesian degradation modelling, both for spare parts inventory management.

4.6. Data driven approaches and machine learning

[Beutel and Minner \(2012\)](#) take a data-driven approach that sets the inventory level as a dependent variable in a linear regression using various explanatory variables of demand (covariates or features). This approach has been adopted by a number of researchers who all report savings over classical inventory control methods (see, e.g., [Huang and van Mieghem, 2014](#); [Shi et al., 2016](#); [Huber et al., 2019](#)).

[Van Steenbergen and Mes \(2020\)](#) use machine learning to forecast the demand of new products (within 18 weeks of introduction) utilising product characteristics of old comparable products. They find improvements in terms of forecast accuracy and inventory costs against a benchmark consisting of an empirical distribution drawn straight from the test data, and another based on the total demand of the most similar product, multiplied by some fixed coefficient of variation.

[Huber et al. \(2019\)](#) use machine learning setups based on linear regression, artificial neural networks (see [Hornik, 1991](#)) and gradient-boosted decision trees (see [Friedman, 2001](#)) to predict

daily demand of a German bakery chain (newsvendor problem). They compare forecast accuracy and cost against a well selected number of benchmarks (including various exponential smoothing methods - ES) and find that the machine learning methods outperform in both accuracy and cost, but only when trained on the entire dataset (ES perform best when the methods are trained on one time series at a time).

[Ban and Rudin \(2019\)](#) apply machine learning algorithms based on the empirical risk minimisation (ERM) principle and kernel optimisation to a single time series of emergency room nurse staffing levels (approximated as a newsvendor problem). They calculate means and 95% confidence intervals, and benchmark against a number of techniques and report improvements against the 'best practice' (naïve seasonal forecast: average demand per day of week). They demonstrate how to carry out a careful data-driven investigation, conclude that there is no single approach to solving the 'big data' newsvendor problem (newsvendor formulation which is solved with help of explanatory variables), and finally warn about the dangers of overfitting.

A main concern for these methods is the fact that they are a 'black box', which means that there is difficulty in justifying the resulting predictions. When improvements in forecasting accuracy can be convincingly showcased against robust, vigorous benchmarks, this is less of an issue (e.g., [Huber et al., 2019](#)). However, this rigorous testing is perhaps not as common as it should be, with machine learning methods often tested in violation of best practice forecasting accuracy testing ([Fildes et al., 2020](#); see [Tashman, 2000](#), for a review of established guidelines). [Spiliotis et al. \(2020\)](#) and [Ma and Fildes \(2020\)](#) offer examples of rigorous testing of new data-driven approaches and show promising results in terms of forecasting accuracy. It would be interesting to see how these improvements translate in inventory savings. Beyond the correct selection of methods to benchmark against, we also see applications on too few SKUs (e.g., [Cao and Shen, 2019](#); [Ban and Rudin, 2019](#)). One reason could be the computational intensity of these methods, although technological advances are making this argument increasingly fragile. [Babai et al. \(2020\)](#) compare the neural network of [Gutierrez \(2008\)](#) and proposed iterations against bootstrapping, simple exponential smoothing and Croston variants, in a dataset of 5135 intermittent demand time-series. They find that simple exponential smoothing outperforms all other methods both in terms of forecast accuracy and inventory performance.

It has also been noted that ML methods do not readily provide predictive densities ([Fildes et al., 2020](#)). There have been however some adaptations to provide probabilistic forecasts. [Wen et al. \(2017\)](#) and [Gasthaus et al. \(2019\)](#) propose using monotonic regression splines (see [Wegman and Wright, 1983](#)) optimised by a neural network with a continuous ranked probability score (see [Section 4.7](#)) objective. [Salinas et al. \(2020\)](#) propose an autoregressive recurrent neural network model for producing probabilistic forecasts. [Van Steenbergen and Mes \(2020\)](#) combine k-means clustering, random forest (see [Breiman, 2001](#)) and quantile regression forest in a machine learning algorithm to compute quantiles and prediction intervals.

4.7. Forecast evaluation

The importance of forecast evaluation cannot be stressed enough, especially when used to select forecasts that are then sequentially informing inventory decisions. Point forecasts are judged on a number of forecasting accuracy metrics (see [Section 3.2.2](#)), and the lack of consensus on what these metrics should be is well discussed in the literature (see, e.g., [Makridakis et al., 2020](#)). A complicating factor is that different point forecasts are optimised at different values depending on the accuracy metric un-

der consideration (Gneiting, 2011a; Kolassa, 2019). For example, absolute errors are consistent for median point forecasts, while squared errors are consistent for the mean (Gneiting and Katzfuss, 2014).

Feeding forecasts into reasonable inventory models is a good way to gain extra confidence on the adequacy of the inventory forecasting assumptions (e.g., normality of residuals). In effect, what is tested then is the veracity of the distributional assumption of the errors and the accuracy of the quantile extrapolation (e.g., from the point forecast of mean demand to the often cycle service level prescribed quantile of interest). What would perhaps be even better, would be to optimise the forecasts directly on the end goal as it is measured through the relevant inventory policy each time in question (Tratar, 2010; Kourentzes et al., 2020).

A more straightforward way to go about this would be to directly judge the forecasting accuracy of the percentile of interest (Gneiting, 2011b). For example, asymmetric piecewise linear loss functions (also known as: the pinball, the linlin, hinge, tick, and newsvendor loss) can be used to compare the relative accuracy of two quantile forecasting models (Koenker, 2005; Gneiting, 2011b). Very often, inventory costs are compared at a number of target cycle service levels, e.g., 90%, 95%, 99%. Such measures can be used to directly test losses at those quantiles of interest.

A growing number of researchers argue that the focus should move from point forecasts (be it mean or some quantile) to forecast full predictive densities (e.g., Gneiting, 2011b; Snyder et al., 2012; Kolassa, 2016; Fildes et al., 2019; de Kok, 2019). Equally important then is to be able to judge the accuracy of these forecasted distributions. But what does accuracy mean in this probabilistic forecasting context? Gneiting et al. (2007) define it as a combination of calibration (statistical consistency between distributional forecasts and observations) and sharpness (concentration of the predictive distributions). Recently, (proper) scoring rules have been introduced that can rank predictive distributions on both these traits (see Gneiting and Raftery, 2007; Gneiting and Katzfus, 2014).

We mention a few scoring rules here for the interested reader, without going into much detail. One such measure is the Continuous Ranked Probability Score (CRPS) (Brown, 1974; Matheson and Winkler, 1976; Gneiting and Raftery, 2007), or its Discrete equivalent (DRPS) (Epstein, 1969; Murphy, 1971; Snyder et al., 2012), with the intuitive definition of a pinball loss across all quantile levels (Gasthaus et al., 2019). Another is the Brier or quadratic score (Brier, 1950). As Boylan and Syntetos (2006) point out however, in most applications of inventory control, we are more interested in particular parts of the predictive distribution (e.g., at 90%+ to correspond with the most common target cycle service levels). The scoring rules offer little guidance as to how competing forecasting methods might fair in those particular percentiles (e.g., the overall winning distribution might be overperforming in parts of the distribution we are not interested in and losing in the parts that we are).

The Probability Integral Transform for continuous density forecasts (PIT) (Rosenblatt, 1952) and the randomised PIT for discrete predictive densities (rPIT) (see, e.g., Kolassa, 2016, and references therein) is another standard way to evaluate distributions. By creating a histogram of PIT values and checking for its uniformity (through goodness-of-fit tests, see, e.g., Inglot and Ledwina, 2006), we can see in which percentiles the predictive distributions differ. In other words, having uniform PITs indicates a well-calibrated forecast; however, it does not inform us of its sharpness (Gneiting et al., 2007). To overcome the deficiencies of proper scoring rules and PIT, Kolassa (2016) proposes these measures to be used in conjunction. Another option is to incorporate quantile-weighted CRPS in alignment with the target cycle service levels of interest (Gneiting and Ranjan, 2011).

Conclusion

We have reviewed the literature of integrated inventory control and forecasting. To do so, we consulted experts in both fields to inform our database keyword search but also the integration framework. The framework defined four levels (0 to 3; see Appendix B). The first two levels describe non-integrated and the last two, which are the focus of our review, integrated approaches. We find the integrated literature started formulating into a stream in the early 1970s. This is logical as integration cannot happen in a vacuum and must be preceded by an in-depth analysis of its constituents (integration levels 0 and 1).

Since then, a growth has been identified, which seems to have plateaued at about 10 articles per year over the last decade. This may be interpreted as a change of focus of authors back to the individual streams, which underlines the importance of this work (as well as others' calls for integration). The analysis of the integrated (levels 2 and 3) sample revealed promising research streams which were followed up with further exploration of the individual streams (including but not constrained to the papers found in our sample). It is worth noting the role of research on slow/intermittent demand which historically approached forecasting and inventory control jointly. The same is also true for research on circular (closed) loops and returns forecasting, which while mostly employing integrated approaches is still an area wide open to contributions.

Historically, forecasters were interested in the performance of (mostly point) forecasts (of mean demand), measured through accuracy metrics serving as proxies for forecasting utility. Inventory controllers, on the other hand, considered forecasts (when they at all did) as an exogenous variable beyond their control or interest, a readily available and to specification input to the inventory control process. The underlying notion is that an expert forecaster would create a 'perfect' forecast (that will accurately describe a true demand distribution as needed), which would then be serially picked up by an expert stockist, to be transformed in as good as possible inventory quantities of interest. Where these forecasts are good (in terms of both accuracy and conformity to inventory control assumptions), literature at levels 0 and 1 (e.g., EOQ formulations) provide optimal inventory decisions.

From the inventory perspective, which for the literature under consideration is in effect the end goal of inventory forecasting, three historical trends have emerged. First was the assumption that every observation is independent and identically distributed, drawn from a real underlying distribution. Using these observations, an estimation of the latter is made, and substituted in its place (and treated as if it was the real distribution). The realisation that demands are more often than not time-correlated led to an adaptation of this process, and to the emergence of a second trend.

Researchers would typically train point forecasts of the mean, assume some distribution of residuals (most often Gaussian with mean of zero), calculate a variance metric (often the mean squared error) and follow tables or algorithms to reach the target service level prescribed quantile. We have seen that this approach is not ideal as a) we tend to judge forecast accuracy for the mean or quantiles irrelevant to our end goals, b) the very convenient normality (or other) distributional assumption of errors is often in violation (Koenker and Basset, 1978).

The third stream then relaxes the assumption of normality, or any distributional or parametric assumption, to move directly from the observations to the quantiles of interest. We outlined examples of this stream in Section 3 and further elaborated on them in Section 4. Bootstrapping, robust optimisation, and density forecasts or quantile estimation are examples of streams attempting just that. Heuristic and increasingly more so machine learning ap-

proaches are called upon to provide solutions to these complex problems. Data driven solution approaches have taken advantage of recent developments in computing power and big data to help with solving them.

A parallel argument that arises from these historic paradigm shifts, or perhaps can be used to explain them, is that of the speed of change. The assumption that demand characteristics remain stable for a long time was perhaps reasonable up to the end of the 20th century, but is maybe not much so anymore (Bowersox, 2007). For one, product life cycles have since reduced, which, other than directly violating the above assumption, also leads to a great reduction in time length of available data (Basallo-Triana et al. 2017, Baardman et al., 2018, Van Steenberghe and Mes, 2020).

While bootstrapping and machine learning approaches do not rely on distributional assumptions, problems arise especially when few demand observations are available. Overfitting and their inaptitude to anticipate possible future values that are not encountered in the sample are the main problems. On the other hand, Bayesian approaches do not face these problems, they do however depend on distributional assumptions.

It is worth noting that the vast majority of research is dedicated to linear supply chains: from resource extraction to serving customer demand. Increasingly however, governments, customers and companies are interested in how to retain resources in circulation (the circular economy) as much as possible before disposal (Goltsov et al., 2019a). We note that demand inventory forecasting should be in this sense expanded to demand and returns inventory forecasting, an area which is open to contributions.

5.1. The role of integrated inventory forecasting

The need for integration has been to an extent substantiated from the combination of levels 0 and 1 research (in level 2 papers, often in exploration of the validity of relevant theoretical assumptions through simulation in real data). However, what this also implicitly highlights is the fact that integration is not required when the distributional assumptions are robust. Integration should be treated as an injection of realism, to treat (or investigate potential) violated assumptions – but not to complicate for the sake of complication. So, literature of levels 0 and 1 provide a mosaic from where to draw solutions when the relevant assumptions (more or less) hold, but also foundations or inspiration to enable and educate the naturally more complex integrated approaches.

Literature at level 2 then seems to identify cases where sequential application of non-integrated approaches underperform. Simulation has played a prominent role in this domain, in particular when combined with industry data, to expose dubious assumptions that lead to performance losses. Literature at level 3 delves deeper and explores the reasons behind the inaccuracies, employing integration as a remedy. Often, non-integrated literature provides bounds for their integrated counterparts. For example, the classical EOQ cost function provides upper cost boundaries for the more realistic, optimal discrete EOQ, with Lagodimos et al. (2018, pp. 119) advising “extreme caution when transferring results between the continuous and discrete-time frameworks” – or in other words, when in violation of the continuous time assumption. The above discussion is graphically summarised in Figure 16.

5.2. The trade-off between complexity and efficiency gains

As the scope widens to accommodate this integration, the analytical and computational burdens increase. Demand assumptions are relaxed, forecasts need to be produced and inventory policies need to accommodate the fact that demand is forecasted, increasing the complexity of mathematical or algorithmic models or sim-

ulations. For these reasons, machine learning techniques, heuristics and simulation are often relied upon to explore the intertwined inventory forecasting problems (solutions).

We have shown many examples where integration has provided better results than non-integrated approaches (see Sections 3 and 4). However, there exists a trade-off between increasing complexity and potential gains from pursuing it. An important aspect of complexity in research is how it affects its adoption by practice. Niemi et al. (2009), investigating the innovation-adoption gap with regard to inventory management techniques, note that “despite all the theory available, the inventory management techniques in use in companies are often very elementary”. The logical question that arises is, when is it worth to address these extended complexities? It follows that integration is not to be pursued for the sake of integration – in other words a case needs to be made in terms of efficiency gains over effort expended.

What is perhaps missing is an evaluation of this trade-off between benefits of integration and the realism of the assumptions, as well as the severity and likelihood of finding oneself in violation. Quite a few papers have attempted to quantify this by benchmarking inventory performance of integrated approaches against more traditional ones (e.g., Taylor, 2007, for the order-up-to level calculation). This should perhaps become the standard, in line with the mostly adhered notion of benchmarking against (also) simple forecasting procedures (such as simple exponential smoothing) when comparing forecasting methods (e.g., Eaves and Kingsman, 2004). (Relatively) simple integration can reveal the scope for improvement (relative to non-integrated approaches) and provide potential areas where deeper exploration would be merited.

What is a simple way to integrate? Integration is a spectrum, and often the distinction between levels 2 and 3 are not clear. We can however attempt to define where it starts. On the lower end (of level 2), the minimum requirement would be a sequential application of forecasting and inventory control. Demand history is analysed, inventory quantities of interest are forecasted, and these forecasts inform an inventory policy. So, from a forecasting perspective, forecasts need to be subject to an analysis of their ‘utility’. From an inventory control perspective, this means that modelling should not overly rely on convenient assumptions of known demand but expose itself to more realistic assumptions of demand and forecasts.

5.3. Forecasting for inventory control

What to forecast (and how to assess it) remains an open, actively researched question. An unbiased forecast of e.g., the median (where the absolute errors optimise in) describes the level of inventory that would satisfy demand (from stock on hand) 50% of the time. The question is, why are we forecasting (and comparing performance against) the median (or mean), if the target service levels are rarely 50% (i.e., in a newsvendor type setting the inventory holding and backorder cost being equal)? That is to say, point (mean/median) forecasts are not enough for inventory control purposes, however accurate they might be.

The utility of the forecasts is in effect a proxy evaluation of how good the inventory forecasting system in consideration is, in predicting a quantile of interest (that directly or indirectly corresponds to a service level; Gardner, 1990). At a minimum, there should also be a forecast for the variance of the forecast errors, perhaps through some mean squared error smoothing procedure (Brown, 1982; Bretschneider, 1986; see Babai et al., 2021b, for a comparison of approaches). But still a (point) forecast of mean demand and variance also requires a distributional assumption to compute safety stocks (i.e., quantiles of interest), and at times re-

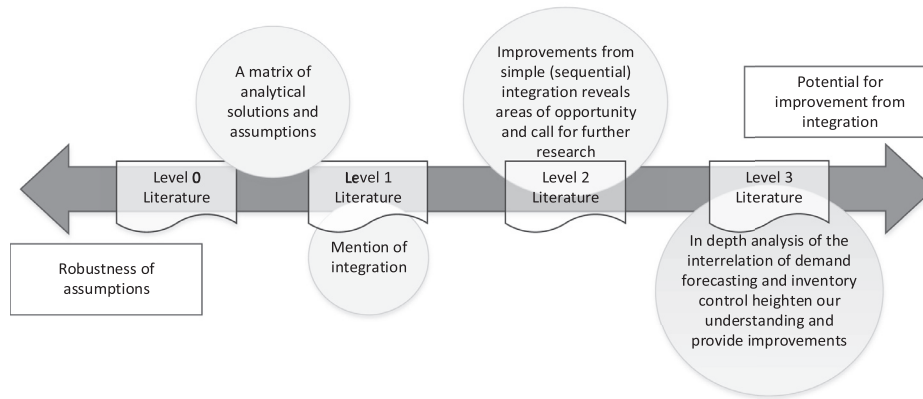


Fig. 16. Integration serves to identify and treat cases where the assumptions are dubious. Integration comes at a cost of complexity, but also with potential for improvements.

lying on the first two moments alone is not enough (Lagodimos et al., 1995).

It seems that one of the most promising approaches, from an integrated inventory forecasting perspective at least, is to forecast the entire lead time demand distribution, bypassing the need to rely on at times dubious distributional assumptions (Taylor, 2007; Snyder et al., 2012; Barrow and Kourentzes, 2016; Kolassa, 2016; Amrani and Khmelnsky, 2017) and then extracting the relevant quantiles – an area in need of further exploration (Fildes et al., 2020). Among the same lines, point forecast accuracy research is progressing (e.g., Petropoulos and Kourentzes, 2015), but the need for accuracy measures for the entire distribution is pressing (Kolassa, 2016).

Bayesian methodologies offer a natural connection to classic inventory theory while using a nimbler approach that updates the parameters of the distribution as more data become available. They readily provide probability densities rather than point forecasts; however, the inventory forecaster still needs to assume a distribution (and the selection can complicate the parametrisation efforts). Robust optimisation and bootstrapping approaches are interesting ways to avoid problematic distributional assumptions, especially when avoiding assumptions of i.i.d. demand. Machine learning approaches show great promise with their ability to calculate inventory parameters of interest directly from the data, while incorporating covariate information (features) beyond the timeseries. More effort should be expended in dealing with issues of overfitting and to employ proper benchmarking.

As we move towards more integrated research on inventory forecasting, the good practice of forecast evaluation should be expanded to include rigorous benchmarking of the entire inventory operation. Such a benchmark should include a relevant inventory policy, with a carefully selected number of appropriate forecasting methods and distributional assumptions. Any parameter (including, e.g., weights for the exponential smoothing family) should be optimised on the bottom-line inventory performance metric (most often, some cost equation alongside a service level). Cross-validation and testing across a number of SKUs are well established aspects of forecasting benchmarking that should be retained here too.

Finally, and perhaps most importantly, this integrated thinking should instigate fundamental, simple solutions to the joint question of inventory forecasting. Beyond any future developments in the streams discussed above, there is certainly tremendous scope in uniting inventory and forecasting by a better conceptual understanding of their interaction. Such theoretic, often closed form formulations, provide solutions that can also assist heuristic, machine learning and other applications by providing solid foundations and freeing up processing power.

Funding

Engineering and Physical Sciences Research Council (EPSRC): project EP/P008925/1; Innovate UK and EPSRC: project KTP10171.

Acknowledgements

We sincerely appreciate discussions and feedback received during the following events, which helped shape our paper: an International Institute of Forecasters workshop in Lancaster (2016); a research seminar in the Technical University of Darmstadt (2016); the International Symposium on Inventories in Budapest (2018); and the International Symposium on Forecasting in Santander (2016). We would also like to acknowledge the contribution of the two anonymous referees, who greatly helped to improve the content of the paper and its presentation.

Appendix A. Search string

Our keyword sets were informed in three ways: i) initial keyword selection and search, ii) (key)word analysis on the resulting paper (pre-)sample, and iii) expert consultation. To initiate the survey of the literature, we created an initial keyword set that returned a pre-sample of the literature (1500 articles). We then went through all titles of the pre-sample papers, and when needed abstracts, excluding irrelevant ones, ending up with 60 articles. Forward and backward searches revealed a further 150 relevant articles, bringing the pre-sample to 210 articles. We performed a content analysis on this pre-sample and produced lists of forecasting- and inventory control-related keywords ranked by instances of appearance in titles, abstracts or keywords⁸. At the same time, we contacted leading academics in the broader fields of forecasting and inventory control, asking them to produce five keywords for each group. The results of both exercises served to educate the final keyword selection.

The selected keywords were thematically split between forecasting and inventory control, and then further into “area-defining” and “context-specific” sets. We put broad keywords that tend to create many hits (of a broad scope) in the former, and the rest, which form an inclusive (albeit not exhaustive) attempt to capture different areas and niches of the relevant literature, in the latter. The ‘context specific’ keyword sets attempt to exclude irrelevant

⁸ Some keywords were excluded for being too broad or of little relevance to our needs (e.g., ‘production’ or ‘observed’), and some were conjoined (‘supply’ and ‘chain’ where the two most frequent words after ‘demand’) or manipulated into a single entry (‘season’, ‘seasonal’ and ‘seasonality’ are represented by ‘seasonal’).

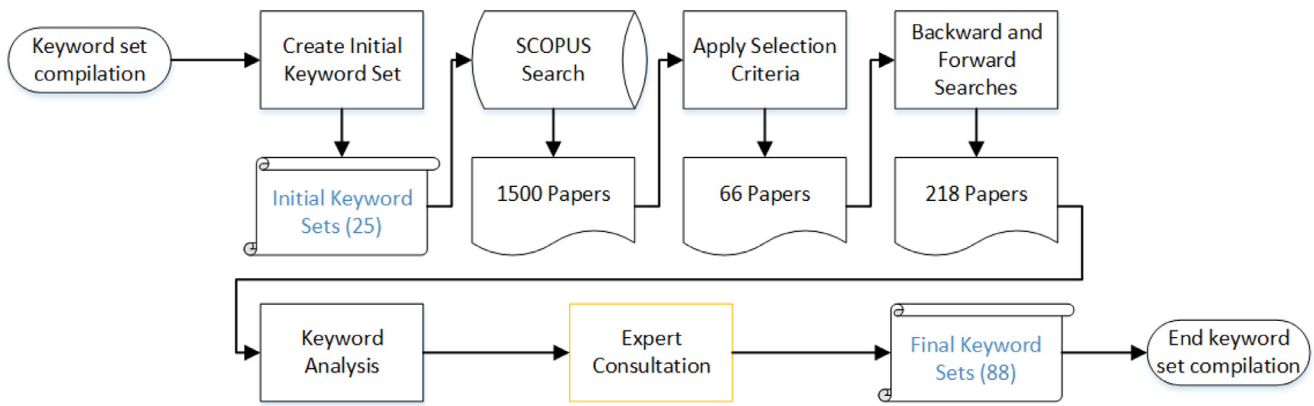


Fig. 17. Keyword sets compilation.

(to our paper) literature that may share some of the keywords (e.g., weather or energy forecasting, or financial/stock market research).

For a paper to be included in the sample, it had to contain at least one word of each group of keywords in its title, abstract or author selected keywords (see Figure 4). We focused on academic papers and restricted ourselves to English manuscripts to ensure readability across the authors. Finally, papers from various areas irrelevant to our search academic fields (e.g., medicine, meteorology, etc. – as these were defined by Scopus) were excluded to help focus the sample.

Appendix B. Levels of integration

Any papers irrelevant to our focus we find in our final sample are considered erroneous inclusions. This in a sense also includes literature of the non-integrated levels 0 and 1 (that may end up in the final sample), as by merit of the search string, we are looking for papers dealing with inventory and forecasting concurrently (and therefore literature of levels 0 and 1 should be largely excluded). While our focus is on the integrated levels, with examples mostly taken from our sample, levels 0 and 1 are included in the framework for completeness, with examples mostly taken from seminal books and reviews of the field. Any integrated literature (levels 2 and 3) that inadvertently ends up outside the sample are considered erroneous exclusions (Fig. 17, Fig. 21).

B1. Integration level 0

This level consists of what may be termed the ‘traditional’ literature on forecasting and inventory control, respectively. It can be seen as a natural first step whereby the research is relying on con-

venient assumptions to solve the simplest possible models (where the research question still exists).

Inventory level 0 models assume that all information on demand (generating process, DGP) that can be obtained at the point of decision making is given, and they do concern themselves with how the information is obtained (IO, Figure 18). There is no mention of any forecasting or estimation of parameters taking place, not even as a recognition that there is a need to do so. Demand is assumed to be known deterministically (e.g., fixed demand rate per unit time as assumed in the basic EOQ formulation in Harris, 1913) or stochastically (e.g., normal distribution with known mean and variance in other formulations in Lagodimos, 1992). Examples of these works can be found in all standard books in operations and supply chain management (e.g., Hadley and Whitin, 1963; Zipkin, 2000; Muckstadt and Sapra, 2010).

Along the same lines, the traditional forecasting literature, where methods’ development and/or testing solely emphasises forecast accuracy, has no mention of any subsequent use of these forecasts. That is to say, in F0 (Figure 19) forecasting is treated as being an end in itself (e.g., Makridakis et al., 1998), not concerning itself with the utility of the forecasts.

Please note that what we represent here as “forecasting accuracy”, “service level” and “inventory cost” can take many forms. Forecast accuracy can mean a variety of things such as point forecast error measurements, e.g., mean squared or absolute errors. Service level might refer to cycle service level or fill rate. Inventory (related) cost can refer to holding costs, ordering costs, backorder costs or costs lost sales, among others. In any case, what we intend to highlight here is that forecasting and/or inventory performance is somehow evaluated. We discuss in Section 3 examples of the measures we encountered in our sample for both forecasting and inventory control.

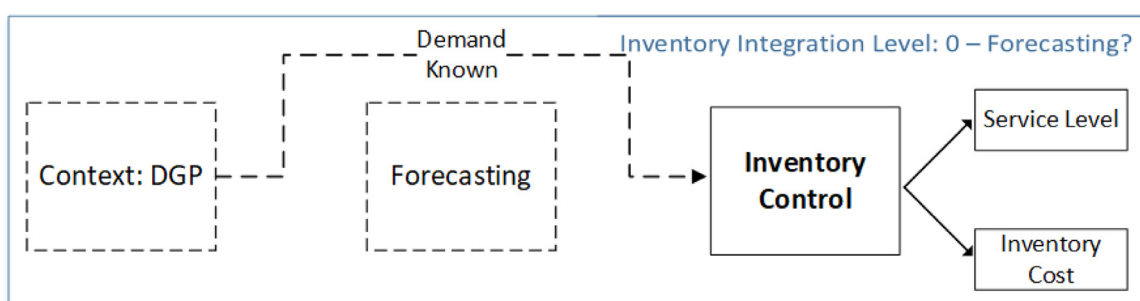


Figure 18. Inventory control level 0 - no mention of forecasting and need to estimate parameters.

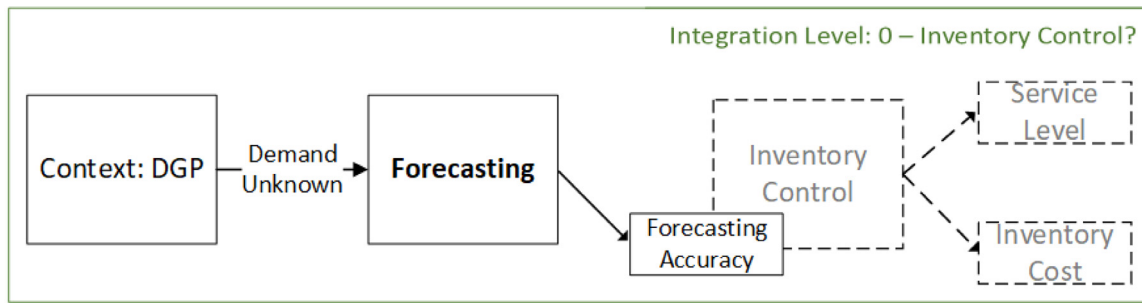


Fig. 19. Forecasting level 0 - no mention of inventory control or any subsequent stages of computation and evaluation.

B2. Integration level 1

This level describes literature very similar to that included in level 0, with one important point of departure: the recognition that demand should actually be forecasted, or that the forecasts are to be used for the end goal of controlling inventories. While these are important qualifications to position the research in the broader context of inventory forecasting, the literature here has otherwise similar modelling decisions and assumptions as it does in level 0.

Research of Inventory level 1 (I1, Figure 20) mentions this need, but forecasting is not employed per se. It is discussed as a separate entity (see Silver et al., 1998, 2017) and with no integration of forecasts (and their errors) in the inventory policies (see Waters, 2008).

Similarly, research in F1 (Figure 23) does not consider forecasting utility metrics or bottom-line inventory implications (i.e., the forecast accuracy implications). That is to say, they do not pursue the investigation of how forecasts are affecting the ultimate goal of controlling inventories, not going much further than mentioning potential inventory implications (e.g., Hyndman and Athanasopoulos, 2018). This is also reflected in seminal reviews of time series forecasting, e.g., see Gardner (1985, 2006) (with focus on exponential smoothing) and De Gooijer and Hyndman (2006).

B3. Integration level 2

At integration level 2, demand is assumed stochastic and unknown, and is forecasted. An inventory policy is turning the forecasts into inventory quantities of interest. The applications at this level are serial in nature, in the sense that forecasting methods and inventory policies are selected (and/or optimised) in isolation, and their interrelation is not closely examined (e.g., Willemain et al., 1994; Eaves and Kingsman, 2004).

Level I2 (Figure 22) describes articles where a forecasting procedure is selected and applied, as an input to an inventory control model, which is where the focus of the paper in question lies. Syntetos and Boylan (2006), for example, investigate the inventory performance (holding cost and fill rate – the percentage of demand served directly from stock on hand) by simulating a periodic (review period T) order-up-to (level S), (T, S) policy, with various forecasting methods as input. Another example is Watson (1987), that incorporated periodic exponential smoothing and moving average forecasts in an EOQ-type model, to explore the effect of forecast errors in achieving service levels.

Similarly, but when the scope and focus of the paper is on forecasting, the paper is assigned the level F2 (Figure 23). Some inven-

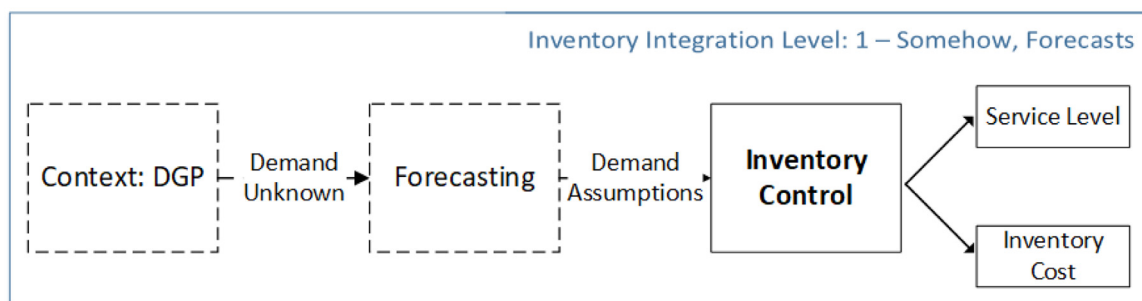


Fig. 20. Inventory level 1 – Same as I0 but with a discussion of a need to "somehow" forecast demand.

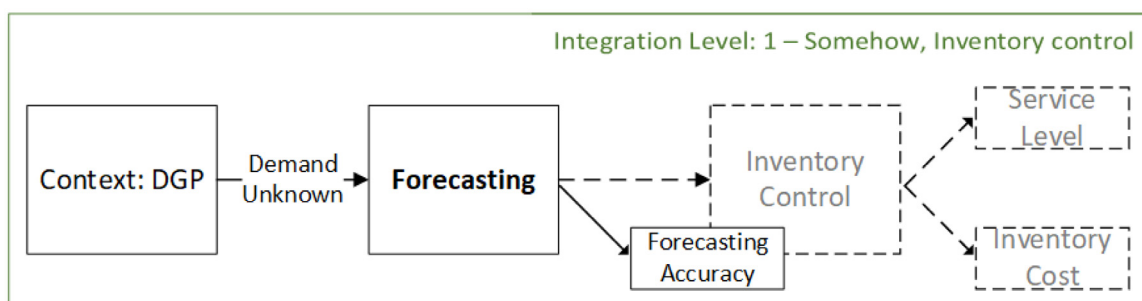


Fig. 21. Forecasting Level 1 – Same as F0 but with a discussion of inventory and potential implications.

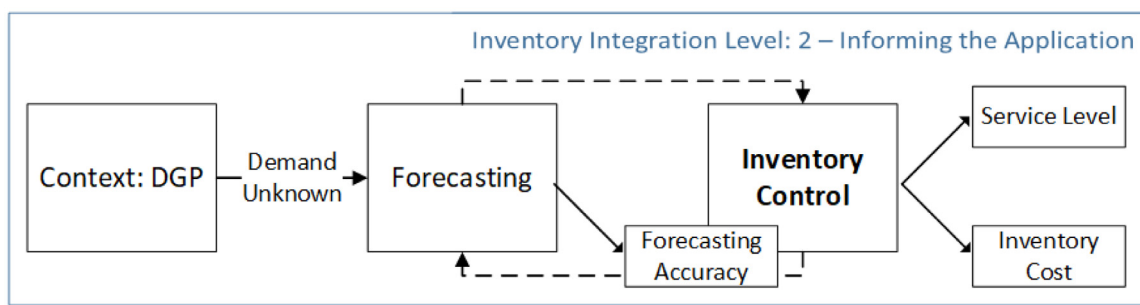


Fig. 22. Inventory control level 2 – Demand is forecasted and is serially informing the inventory control process.

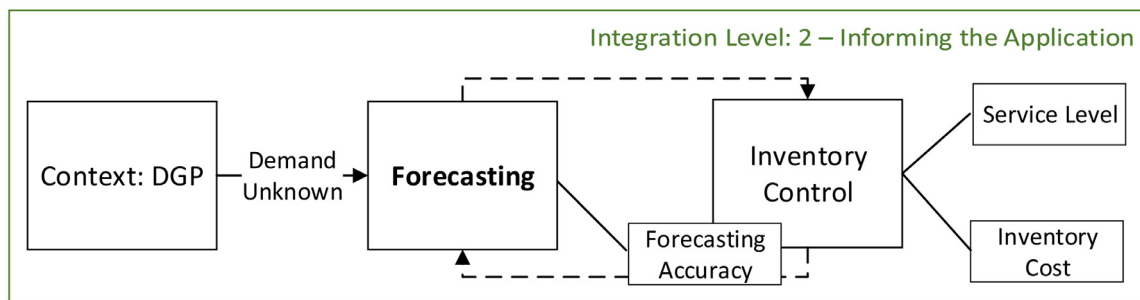


Fig. 23. Forecasting level 2 – Demand is forecasted and is serially informing an inventory control process.

tory policy is employed, and there is some evaluation of the forecast's 'utility' in terms of inventory implications (most often cost and service level). Gardner (1990) makes a very good example, employing efficiency curves⁹ between different forecasting procedures to evaluate their inventory performance in a reorder point (r) order quantity (Q), (r, Q) inventory policy.

B4. Integration level 3

Integration level 3 is reserved for papers that approach the problem of inventory forecasting jointly. That is to say, they take an integrated approach whereby some contribution is achieved by looking at the entire picture. In that sense, forecasting modelling decisions are influenced by the inventory modelling decisions (and vice versa), and most often a joint metric is pursued. A holistic understanding of the specific (and joint) nature of the inventory forecasting problem is required as it is furthered. This would then constitute a development towards inventory forecasting as a joint entity. As such, the framework converges at level IF3, to describe integrated inventory forecasting research. Literature here goes beyond serial application and is more often than not concerned with the interrelations of forecasting and inventory control in specific demand contexts in violation of common (convenient) assumptions. At a minimum, modelling decisions relating to both forecasting and inventory control are judged on inventory performance metrics (e.g., optimizing simple exponential smoothing parameter alpha directly on inventory performance in Kourentzes et al., 2020).

The fact that demand parameters are forecasted has its own implications, and literature at IF3 (Figure 24) attempts to investigate and address them. For example, Prak et al. (2017) observed that while demand may not be correlated, forecast errors are, leading to undershoots when neglected. They proposed appropriate safety stock adjustment mechanisms when the one-step-ahead

forecast errors are informing the variance calculations. Hoberg et al. (2007) compared a linear and a proposed non-linear (integrated) inventory policy with simple exponential smoothing forecasts, against stationary and non-stationary demand, and found the integrated approach reduces order amplification. Dejonckheere et al. (2002) employed transfer functions to investigate the influence of forecast errors in the bullwhip effect (inventory variance over demand variance) in supply chains. The above are a collection of papers showing how the research interests in IF3 have grown to include the interrelations of forecasting and inventory control.

Forecasts are moving away from forecasting the mean to also forecasting the variance (Brown, 1962; Bretschneider, 1986; Snyder, 2004), but also further into forecasting the entire lead time distribution (Barrow and Kourentzes, 2016). In the same wavelength, effort is expended to propose forecasting accuracy metrics that better represent the end goal of inventory control, including ways to judge forecasts in their ability to forecast percentiles of interest (Kolassa, 2016). While the performance of the system will be ultimately judged by appropriate inventory metrics, there is still merit to track forecasting performance as it can help to identify issues in the forecasting process – and track performance losses to the demand generating process.

Finally, we note that the proposed integration levels attempt to classify a semi-abstract spectrum of integrated literature. As such, there is a certain degree of subjectivity in assigning the levels for each paper. Especially when it comes to books, certain chapters might be of different levels than others (e.g., Silver et al., 1998, and subsequent editions, could be seen as a mostly level F1 book with a chapter on forecasting being level F2).

B.5. Classification beyond integration

Beyond the assignment of an integration level to each paper (and as a stepping-stone to help us reach that decision), we have recorded a further number of variables. These are presented below.

- Bibliometric information: This information is directly extracted from Scopus. We are interested in the year of publication and

⁹ Curves (efficiency frontiers) contrasting different methods against some inventory performance measures, e.g., cost against achieved service level (Brown, 1967; Gardner and Dannenbring, 1979).

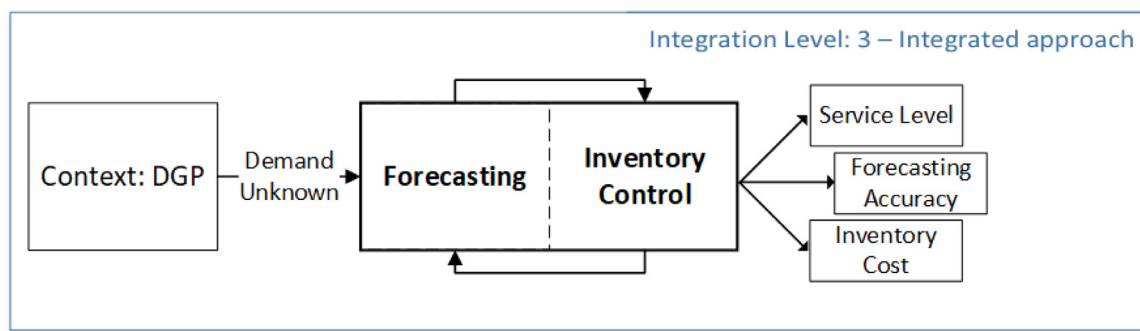


Fig. 24. Inventory forecasting level 3 - Full integration.

journal of publication. This will help us get a feel for the area and its size over time.

- **Methods and policies employed:** We record forecasting methods used to forecast demand (e.g., exponential smoothing family), as well as inventory policies used to satisfy demand (e.g., order-up-to). We also note when particular model (e.g., state space) or approach (e.g., aggregation) or solution methodologies (e.g., heuristics or machine learning) are employed. If a paper employs two or more policies and/or methods (e.g., for comparison purposes), the record will reflect all employed, as required for forecasting or for inventory control.
- **Performance measurement:** Here we capture information on the metrics used for the evaluation of either the forecasts or the inventory performance. These are forecasting accuracy metrics (e.g., mean squared error) and inventory performance metrics (e.g., bullwhip effect, cost, or average inventory), respectively. Again, multiple entries on both forecasting and inventory metrics are captured.
- **Methodology-related information:** We also record whether a paper pursued some theoretical development, making no judgement on veracity or significance. Introducing new formulae, models, algorithms, proofs of lemmas/theorems, are judged as an analytical development. Repetitions of methods proposed in other papers do not qualify, while adaptations do. We similarly note if simulation is taking place on any kind of data and scenarios. Simple numerical examples and simulations that occur elsewhere and are not reported in the paper (omitted) do not qualify. These two variables could simultaneously be true or false for a single paper.
- **Extraneous information:** Further to the above we also record if a paper is dealing with slow/intermittent demand, if it concerns multiple nodes in a supply chain, or closed loop supply chains. Finally, we make a note if data employed (e.g., demand/sales time series) are empirical (e.g., time series taken from industry) or theoretically generated (e.g., when demand is drawn from a normal distribution with e.g., mean $\mu = 200$ and standard deviation $\sigma = 20$).

References

- Amrani, H., & Khmelnitsky, E. (2017). Estimation of quantiles of non-stationary demand distributions. *IIE Transactions*, 49(4), 381–394.
- Andriolo, A., Battini, D., Grubbström, R. W., Persona, A., & Sgarbossa, F. (2014). A century of evolution from Harris' s basic lot size model: Survey and research agenda. *International Journal of Production Economics*, 155, 16–38.
- Arrow, K. J., Karlin, S., & Scarf, H. (1958). *Studies in the mathematical theory of inventory and production*. CA: Stanford University Press.
- Axsäter, S. (2015). *Inventory control*: Vol. 225. Springer.
- Azoury, K. S. (1985). Bayes solution to dynamic inventory models under unknown demand distribution. *Management science*, 31(9), 1150–1160.
- Azoury, K. S., & Miyaoka, J. (2009). Optimal policies and approximations for a Bayesian linear regression inventory model. *Management Science*, 55(5), 813–826.
- Baardman, L., Levin, I., Perakis, G., & Singhvi, D. (2018). *Leveraging comparables for new product sales forecasting* (27, pp. 2340–2343). Production and Operations Management.
- Babai, M. Z., Chen, H., Syntetos, A. A., & Lengu, D. (2021a). A compound-Poisson Bayesian approach for spare parts inventory forecasting. *International Journal of Production Economics*, 232, Article 107954 p.
- Babai, M. Z., Dai, Y., Li, Q., Syntetos, A. A., & Wang, X. (2021b). Forecasting of lead-time demand variance: implications for safety stock calculations. *European Journal of Operational Research*.
- Babai, M. Z., Dallery, Y., Boubaker, S., & Kalai, R. (2019). A new method to forecast intermittent demand in the presence of inventory obsolescence. *International Journal of Production Economics*, 209, 30–41.
- Babai, M. Z., Syntetos, A. A., & Teunter, R. (2010). On the empirical performance of (T,s,S) heuristics. *European Journal of Operational Research*, 202(2), 466–472.
- Babai, M. Z., Syntetos, A. A., & Teunter, R. (2014). Intermittent demand forecasting: An empirical study on accuracy and the risk of obsolescence. *International Journal of Production Economics*, 157, 212–219.
- Babai, M. Z., Tsadiras, A., & Papadopoulos, C. (2020). On the empirical performance of some new neural network methods for forecasting intermittent demand. *IMA Journal of Management Mathematics*, 31(3), 281–305.
- Ban, G. Y., Gallien, J., & Mersereau, A. J. (2019). Dynamic procurement of new products with covariate information: The residual tree method. *Manufacturing & Service Operations Management*, 21(4), 798–815.
- Ban, G. Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1), 90–108.
- Barrow, D. K., & Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: implications for inventory management. *International Journal of Production Economics*, 177, 24–33.
- Basallo-Triana, M. J., Rodríguez-Sarasty, J. A., & Benitez-Restrepo, H. D. (2017). Analogue-based demand forecasting of short life-cycle products: a regression approach and a comprehensive assessment. *International Journal of Production Research*, 55(8), 2336–2350.
- Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). *Robust optimization*. Princeton university press.
- Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, 23(4), 769–805.
- Ben-Tal, A., & Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations research letters*, 25(1), 1–13.
- Ben-Tal, A., & Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3), 411–424.
- Bertsimas, D., Sim, M., & Zhang, M. (2019). Adaptive distributionally robust optimization. *Management Science*, 65(2), 604–618.
- Bertsimas, D., & Sturt, B. (2020). Computation of exact bootstrap confidence intervals: complexity and deterministic algorithms. *Operations Research*, 68(3), 949–964.
- Bertsimas, D., & Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations research*, 54(1), 150–168.
- Beutel, A. L., & Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, 140(2), 637–645.
- Bijvank, M., Huh, W. T., Janakiraman, G., & Kang, W. (2014). Robustness of order-up-to policies in lost-sales inventory systems. *Operations Research*, 62(5), 1040–1047.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307–327.
- Bookbinder, J. H., & Lordahl, A. E. (1989). Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE transactions*, 21(4), 302–312.
- Boutselis, P., & McNaught, K. (2019). Using Bayesian Networks to forecast spares demand from equipment failures in a changing service logistics context. *International Journal of Production Economics*, 209, 325–333.
- Bowersox, D. J. (2007). SCM: The past is prologue. *Supply Chain Quarterly*, 1(1), 28–33.
- Boylan, J. E., & Syntetos, A. A. (2006). Accuracy and accuracy-implication metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4, 39–42.

- Boylan, J. E., & Syntetos, A. A. (2021). *Intermittent demand forecasting: Context, methods and applications*. John Wiley & Sons.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Bretschneider, S. (1986). Estimating forecast variance with exponential smoothing: Some new results. *International Journal of Forecasting*, 2(3), 349–355.
- Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78, 1–3.
- Brown, R. G. (1962). *Smoothing, forecasting and prediction of discrete time series*. Prentice-Hall.
- Brown, R. G. (1967). *Decision rules for inventory management*. Holt, Rinehart & Winston.
- Brown, R.G., 1982. *Advanced service parts inventory control*. Materials management systems.
- Brown, T.A., 1974. Admissible scoring systems for continuous distributions. ERIC. Available at: <http://eric.ed.gov/?id=ED135799>. Accessed on 24 January, 2021.
- Cao, Y., & Shen, Z. J. M. (2019). Quantile forecasting and data-driven inventory management under nonstationary demand. *Operations Research Letters*, 47(6), 465–472.
- Cattani, K. D., Jacobs, F. R., & Schoenfelder, J. (2011). Common inventory modeling assumptions that fall short: Arborescent networks, Poisson demand, and single-echelon approximations. *Journal of Operations Management*, 29(5), 488–499.
- Cervellera, C., & Macciò, D. (2011). A comparison of global and semi-local approximation in T-stage stochastic optimization. *European Journal of Operational Research*, 208(2), 109–118.
- Chatfield, D. C., & Hayya, J. C. (2007). All-zero forecasts for lumpy demand: a factorial study. *International Journal of Production Research*, 45(4), 935–950.
- Chen, L. (2010). Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Operations research*, 58(2), 396–413.
- Clements, M. P., & Taylor, N. (2001). Bootstrapping prediction intervals for autoregressive models. *International Journal of Forecasting*, 17(2), 247–267.
- Clottey, T., Benton, W. C., Jr., & Srivastava, R. (2012). Forecasting product returns for remanufacturing operations. *Decision Sciences*, 43(4), 589–614.
- Conrad, S. A. (1976). Sales data and the estimation of demand. *Journal of the Operational Research Society*, 27(1), 123–127.
- Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, 23(3), 289–303.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510–522.
- de Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, 22(3), 443–473.
- de Kok, S. (2019). A Primer on Probabilistic Demand Planning. *The Journal of Business Forecasting*, 38(4), 24–26.
- de Kok, T. (2018). Inventory management: Modeling real-life supply chains and empirical validity. *Foundations and Trends® in Technology, Information and Operations Management*, 11(4), 343–437.
- de Kok, T., Grob, C., Laumanns, M., Minner, S., Rambau, J., & Schade, K. (2018). A typology and literature review on stochastic multi-echelon inventory models. *European Journal of Operational Research*, 269(3), 955–983.
- Dejonckheere, J., Disney, S. M., Lambrecht, M. R., & Towill, D. R. (2002). Transfer function analysis of forecasting induced bullwhip in supply chains. *International journal of production economics*, 78(2), 133–144.
- Diks, E. B., De Kok, A. G., & Lagodimos, A. G. (1996). Multi-echelon systems: A service measure perspective. *European Journal of Operational Research*, 95(2), 241–263.
- Dvoretzky, A., Kiefer, J., & Wolfowitz, J. (1952). The inventory problem: II. Case of unknown distributions of demand. *Econometrica: Journal of the Econometric Society*, 450–466.
- Eaves, A. H., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society*, 55(4), 431–437.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 1–26.
- El Ghaoui, L., & Lebret, H. (1997). Robust solutions to least-squares problems with uncertain data. *SIAM Journal on matrix analysis and applications*, 18(4), 1035–1064.
- El Ghaoui, L., Oustry, F., & Lebret, H. (1998). Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 9(1), 33–52.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 50(4), 987–1007.
- Eppen, G. D., & Martin, R. K. (1988). Determining safety stock in the presence of stochastic lead time and demand. *Management Science*, 34(11), 1380–1390.
- Eppen, G. D., & Schrage, L. (1981). Centralized ordering policies in a multi-warehouse system with lead times and random demand. *Multi-level production/inventory control systems: Theory and practice* (L. B.Schwarz ed.) (pp. 51–68). North-Holland.
- Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology*, 8(6), 985–987.
- Fagundes, M. V. C., Teles, E. O., Vieira de Melo, S. A., & Freires, F. G. M. (2020). Supply chain risk management modelling: A systematic literature network analysis review. *IMA Journal of Management Mathematics*, 31(4), 387–416.
- Fildes, R., & Beard, C. (1992). Forecasting systems for production and inventory control. *International Journal of Operations & Production Management*, 12(5), 4–27.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting*, 25(1), 3–23.
- Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*.
- Fildes, R., Ma, S., & Kolassa, S. (2020). Retail forecasting: Research and practice. *International Journal of Forecasting*. <https://doi.org/10.1016/j.ijforecast.2019.06.004>.
- Flores, B. E., Olson, D. L., & Pearce, S. L. (1993). Use of cost and accuracy measures in forecasting method selection: a physical distribution example. *The International Journal of Production Research*, 31(1), 139–160.
- Fricker, R. D., Jr., & Goodhart, C. A. (2000). Applying a bootstrap approach for setting reorder points in military supply systems. *Naval Research Logistics (NRL)*, 47(6), 459–478.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.
- Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3), 471–483.
- Gardner, E. S., Jr (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1), 1–28.
- Gardner, E. S. (1990). Evaluating forecast performance in an inventory control system. *Management Science*, 36(4), 490–499.
- Gardner, E. S., Jr (2006). Exponential smoothing: The state of the art—Part II. *International journal of forecasting*, 22(4), 637–666.
- Gardner, E. S., & Dannenbring, D. G. (1979). Using optimal policy surfaces to analyze aggregate inventory tradeoffs. *Management Science*, 25(8), 709–720.
- Gasthaus, J., Benidis, K., Wang, Y., Rangapuram, S. S., Salinas, D., Flunkert, V., & Januschowski, T. (2019). Probabilistic forecasting with spline quantile function RNNs. In *The 22nd international conference on artificial intelligence and statistics* (pp. 1901–1910). PMLR.
- Ghodrati, B., & Kumar, U. (2005). Reliability and operating environment-based spare parts estimation approach. *Journal of Quality in Maintenance Engineering*.
- Glock, C. H., Grosse, E. H., & Ries, J. M. (2014). The lot sizing problem: A tertiary study. *International Journal of Production Economics*, 155, 39–51.
- Gneiting, T. (2011a). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494), 746–762.
- Gneiting, T. (2011b). Quantiles as optimal point forecasts. *International journal of forecasting*, 27(2), 197–207.
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2), 243–268.
- Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1, 125–151.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477), 359–378.
- Gneiting, T., & Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3), 411–422.
- Goltso, T. E., Ponte, B., Wang, S., Liu, Y., Naim, M. M., & Syntetos, A. A. (2019a). The boomerang returns? Accounting for the impact of uncertainties on the dynamics of remanufacturing systems. *International Journal of Production Research*, 57(23), 7361–7394.
- Goltso, T. E., Syntetos, A. A., & van der Laan, E. (2019b). Forecasting for remanufacturing: The effects of serialization. *Journal of Operations Management*, 65(5), 447–467.
- Graves, S. C. (1999). A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management*, 1(1), 50–61.
- Gumus, A. T., Guneri, A. F., & Ulengin, F. (2010). A new methodology for multi-echelon inventory management in stochastic and neuro-fuzzy environments. *International Journal of Production Economics*, 128(1), 248–260.
- Ha, C., Seok, H., & Ok, C. (2018). Evaluation of forecasting methods in aggregate production planning: A cumulative absolute forecast error (CAFE). *Computers & Industrial Engineering*, 118, 329–339.
- Hadley, G. W., & Whitin, T. (1963). *Analysis of inventory systems* TM, 1963. Prentice-Hall.
- Harris, F. W. (1913). How many parts to make at once. *Factory, The Magazine of Management*, 10(2), 135–136 152.
- Hasni, M., Aguir, M. S., Babai, M. Z., & Jemai, Z. (2019a). Spare parts demand forecasting: a review on bootstrapping methods. *International Journal of Production Research*, 57(15–16), 4791–4804.
- Hasni, M., Babai, M. Z., Aguir, M. S., & Jemai, Z. (2019b). An investigation on bootstrapping forecasting methods for intermittent demands. *International Journal of Production Economics*, 209, 20–29.
- Heath, D. C., & Jackson, P. L. (1994). Modeling the evolution of demand forecasts ITH application to safety stock analysis in production/distribution systems. *IIE transactions*, 26(3), 17–30.
- Ho, T. H., Savin, S., & Terwiesch, C. (2002). Managing demand and sales dynamics in new product diffusion under supply constraint. *Management science*, 48(2), 187–206.
- Hoberg, K., Bradley, J. R., & Thonemann, U. W. (2007). Analyzing the effect of the inventory policy on order and inventory variability with linear control theory. *European Journal of Operational Research*, 176(3), 1620–1642.
- Hsieh, M. C., Giloni, A., & Hurvich, C. (2020). The propagation and identification of ARMA demand under simple exponential smoothing: forecasting expertise and information sharing. *IMA Journal of Management Mathematics*, 31(3), 307–344.
- Huang, T., & Van Mieghem, J. A. (2014). *Clickstream data and inventory management: Model and empirical analysis* (23, pp. 333–347). Production and Operations Management.

- Huber, J., Müller, S., Fleischmann, M., & Stuckenschmidt, H. (2019). A data-driven newsvendor problem: From data to decision. *European Journal of Operational Research*, 278(3), 904–915.
- Hyndman, R.J., & Athanasopoulos, G., 2018. *Forecasting: principles and practice*, 2nd edition, OTexts. OTexts.com/fpp2. Accessed on 19 January, 2020.
- Ingot, T., & Ledwina, T. (2006). Towards data driven selection of a penalty function for data driven Neyman tests. *Linear algebra and its applications*, 417(1), 124–133.
- Iyer, A. V., & Schrage, L. E. (1992). Analysis of the deterministic (s, S) inventory problem. *Management Science*, 38(9), 1299–1313.
- Jaipuria, S., & Mahapatra, S. S. (2014). An improved demand forecasting method to reduce bullwhip effect in supply chains. *Expert Systems with Applications*, 41(5), 2395–2408.
- John, S., Naim, M. M., & Towill, D. R. (1994). Dynamic analysis of a WIP compensated decision support system. *International Journal of Manufacturing System Design*, 1(4), 283–297.
- Johnston, F. R., & Harrison, P. J. (1986). The variance of lead-time demand. *Journal of the Operational Research Society*, 37(3), 303–308.
- Johnston, F. R., Shale, E. A., Kapoor, S., True, R., & Sheth, A. (2011). Breadth of range and depth of stock: forecasting and inventory management at Euro Car Parts Ltd. *Journal of the Operational Research Society*, 62(3), 433–441.
- Karmarkar, U. S. (1994). A robust forecasting technique for inventory and leadtime management. *Journal of operations management*, 12(1), 45–54.
- Kim, B. S., & Do Chung, B. (2017). Affinely adjustable robust model for multiperiod production planning under uncertainty. *IEEE Transactions on Engineering Management*, 64(4), 505–514.
- Koenker, R. (2005). Quantile regression. *Econometric Society Monographs*, 38.
- Koenker, R., & Bassett, G., Jr (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 46(1), 33–50.
- Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American statistical association*, 101(475), 980–990.
- Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting*, 32(3), 788–803.
- Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, 225, Article 107597.
- Lagodimos, A. G. (1992). Multi-echelon service models for inventory systems under different rationing policies. *International Journal of Production Research*, 30(4), 939–956.
- Lagodimos, A. G., Christou, I. T., & Skouri, K. (2012). Computing globally optimal (s, S, T) inventory policies. *Omega*, 40(5), 660–671.
- Lagodimos, A. G., De Kok, A. G., & Verrijdt, J. H. C. M. (1995). The robustness of multi-echelon service models under autocorrelated demands. *Journal of the Operational Research Society*, 46(1), 92–103.
- Lagodimos, A. G., Skouri, K., Christou, I. T., & Chountalas, P. T. (2018). The discrete-time EOQ model: Solution and implications. *European Journal of Operational Research*, 266(1), 112–121.
- Lau, H. S., & Lau, A. H. L. (1996). Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research*, 92(2), 254–265.
- Lee, C. Y., & Liang, C. L. (2018). Manufacturer's printing forecast, reprinting decision, and contract design in the educational publishing industry. *Computers & Industrial Engineering*, 125, 678–687.
- Li, Q., Disney, S. M., & Gaalman, G. (2014). Avoiding the bullwhip effect using Damped Trend forecasting and the Order-Up-To replenishment policy. *International Journal of Production Economics*, 149, 3–16.
- Lin, J., Naim, M. M., Purvis, L., & Gosling, J. (2017). The extension and exploitation of the inventory and order based production control system archetype from 1982 to 2015. *International Journal of Production Economics*, 194, 135–152.
- Ma, S., & Fildes, R. (2020). Retail sales forecasting with meta-learning. *European Journal of Operational Research*, 288(1), 111–128.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1), 54–74.
- Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting: Methods and applications*. John Wiley & Sons.
- Mason-Jones, R., & Towill, D. R. (1998). Shrinking the supply chain uncertainty circle. *IOM control*, 24(7), 17–22.
- Matheson, J. E., & Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management science*, 22(10), 1087–1096.
- Miller, R. G. (1974). The jackknife - a review. *Biometrika*, 61(1), 1–15.
- Muckstadt, J. A., & Sapra, A. (2010). *Principle of inventory management*. Springer.
- Mulvey, J. M., Vanderbei, R. J., & Zenios, S. A. (1995). Robust optimization of large-scale systems. *Operations research*, 43(2), 264–281.
- Murphy, A. H. (1971). A note on the ranked probability score. *Journal of Applied Meteorology*, 10(1), 155–156.
- Niemi, P., Huiskonen, J., & Kärkkäinen, H. (2009). Understanding the knowledge accumulation process—Implications for the adoption of inventory management techniques. *International Journal of Production Economics*, 118(1), 160–167.
- Ord, K., & Fildes, R. (2012). *Principles of business forecasting*. Nelson Education.
- Ord, K., Fildes, R., & Kourentzes, N. (2017). *Principles of business forecasting* (2nd ed.). Wessex Press Publishing Co.
- Oroojlooyjadid, A., Snyder, L. V., & Takáč, M. (2020). Applying deep learning to the newsvendor problem. *IIE Transactions*, 52(4), 444–463.
- Petropoulos, F., & Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society*, 66(6), 914–924.
- Prak, D., & Teunter, R. (2019). A general method for addressing forecasting uncertainty in inventory models. *International Journal of Forecasting*, 35(1), 224–238.
- Prak, D., Teunter, R., Babai, M. Z., Boylan, J. E., & Syntetos, A. (2021). Robust compound Poisson parameter estimation for inventory control. *Omega*, 104, Article 102481.
- Prak, D., Teunter, R., & Syntetos, A. (2017). On the calculation of safety stocks when demand is forecasted. *European Journal of Operational Research*, 256(2), 454–461.
- Prestwich, S. D., Tarim, S. A., Rossi, R., & Hnich, B. (2014). Forecasting intermittent demand by hyperbolic-exponential smoothing. *International Journal of Forecasting*, 30(4), 928–933.
- Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 11(1), 68–84.
- Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4), 353–360.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of mathematical statistics*, 23(3), 470–472.
- Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of statistics*, 9(1), 130–134.
- Ruiz, C., Pohl, E., & Liao, H. (2021). Bayesian degradation modelling for spare parts inventory management. *IMA Journal of Management Mathematics*, 32(1), 31–49.
- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3), 1181–1191.
- Sani, B., & Kingsman, B. G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 48(7), 700–713.
- Scarf, H. (1959). Bayes solutions of the statistical inventory problem. *The Annals of Mathematical Statistics*, 30(2), 490–508.
- Schneider, H. (1981). Effect of service-levels on order-points or order-levels in inventory models. *The International Journal of Production Research*, 19(6), 615–631.
- See, C. T., & Sim, M. (2010). Robust approximation to multiperiod inventory management. *Operations Research*, 58(3), 583–594.
- Shi, C., Chen, W., & Duenyas, I. (2016). Nonparametric data-driven algorithms for multiproduct inventory systems with censored demand. *Operations Research*, 64(2), 362–370.
- Sillanpää, V., & Liesiö, J. (2018). Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions. *International Journal of Production Research*, 56(12), 4168–4185.
- Silver, E. A., Pyke, D. F., & Peterson, R. (1998). *Inventory management and production planning and scheduling* (3rd edition). Wiley.
- Silver, E. A., Pyke, D. F., & Thomas, D. J. (2017). *Inventory and production management in supply chains* (4th edition). CRC Press.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*: Vol. 26. CRC press.
- Snyder, R. D., Koehler, A. B., Hyndman, R. J., & Ord, J. K. (2004). Exponential smoothing models: Means and variances for lead-time demand. *European Journal of Operational Research*, 158(2), 444–455.
- Snyder, R. D., Koehler, A. B., & Ord, J. K. (2002). Forecasting for inventory control with exponential smoothing. *International Journal of Forecasting*, 18(1), 5–18.
- Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, 28(2), 485–496.
- Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5), 1154–1157.
- Spiliotis, E., Makridakis, S., Semenoglou, A. A., & Assimakopoulos, V. (2020). Comparison of statistical and machine learning methods for daily SKU demand forecasting. *Operational Research*, 1–25.
- Su, C. T., & Wong, J. T. (2008). Design of a replenishment system for a stochastic dynamic production/forecast lot-sizing problem under bullwhip effect. *Expert Systems with Applications*, 34(1), 173–180.
- Syntetos, A. A. (2016). *Forecasting and inventory control: Mind the gap*. Keynote. In *Practice Stream, International Symposium on Forecasting* June 19–22, 2016.
- Syntetos, A. A., Babai, M. Z., Davies, J., & Stephenson, D. (2010). Forecasting and stock control: A study in a wholesaling context. *International Journal of Production Economics*, 127(1), 103–111.
- Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of forecasting*, 21(2), 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. (2008). Demand forecasting adjustments for service-level achievement. *IMA Journal of Management Mathematics*, 19(2), 175–192.
- Syntetos, A. A., Kholidasari, I., & Naim, M. M. (2016). The effects of integrating management judgement into OUT levels: in or out of context? *European Journal of Operational Research*, 249(3), 853–863.
- Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118(1), 72–81.
- Tan, B., & Karabati, S. (2004). Can the desired service level be achieved when the demand and lost sales are unobserved? *IIE Transactions*, 36(4), 345–358.
- Tang, H., Zhang, H., Liu, R., & Du, Y. (2020). Integrating multi-index materials classification and inventory control in discrete manufacturing industry: Using a hybrid ABC-chaos algorithm. *IEEE Transactions on Engineering Management*.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, 16(4), 437–450.
- Taylor, J. W. (2007). Forecasting daily supermarket sales using exponentially weighted quantile regression. *European Journal of Operational Research*, 178(1), 154–167.

- Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, 214(3), 606–615.
- Thorsen, A., & Yao, T. (2017). Robust inventory control under demand and lead time uncertainty. *Annals of Operations Research*, 257(1), 207–236.
- Tiacci, L., & Saetta, S. (2009). An approach to evaluate the impact of interaction between demand forecasting method and stock control policy on the inventory system performances. *International Journal of Production Economics*, 118(1), 63–71.
- Toktay, L. B., Wein, L. M., & Zenios, S. A. (2000). Inventory management of remanufacturable products. *Management Science*, 46(11), 1412–1426.
- Towill, D. R. (1982). Dynamic analysis of an inventory and order based production control system. *The international journal of production research*, 20(6), 671–687.
- Trapero, J. R., Cardós, M., & Kourentzes, N. (2019a). Empirical safety stock estimation based on kernel and GARCH models. *Omega*, 84, 199–211.
- Trapero, J. R., Cardós, M., & Kourentzes, N. (2019b). Quantile forecast optimal combination to enhance safety stock estimation. *International Journal of Forecasting*, 35(1), 239–250.
- Trapero, J. R., Fildes, R., & Davydenko, A. (2011). Nonlinear identification of judgmental forecasts effects at SKU level. *Journal of Forecasting*, 30(5), 490–508.
- Tratar, L. F. (2010). Joint optimisation of demand forecasting and stock control parameters. *International Journal of Production Economics*, 127(1), 173–179.
- van Steenbergen, R. M., & Mes, M. R. K. (2020). Forecasting demand profiles of new products. *Decision support systems*, 139, Article 113401.
- Van Wingerden, E., Basten, R. J. I., Dekker, R., & Rustenburg, W. D. (2014). More grip on inventory control through improved forecasting: A comparative study at three companies. *International journal of production economics*, 157, 220–237.
- Viswanathan, S., & Zhou, C. X. (2008). A new bootstrapping based method for forecasting and safety stock determination for intermittent demand items. *Nanyang Business School*. Nanyang Technological University Singapore. Working paper.
- Wagner, H. M., & Whitin, T. M. (1958). Dynamic version of the economic lot size model. *Management Science*, 5(1), 89–96.
- Wang, Z., & Mersereau, A. J. (2017). Bayesian inventory management with potential change-points in demand. *Production and Operations Management*, 26(2), 341–359.
- Waters, D. (2008). *Inventory control and management*. John Wiley & Sons.
- Watson, R. B. (1987). The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *Journal of the Operational Research Society*, 38(1), 75–82.
- Wegman, E. J., & Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association*, 78(382), 351–365.
- Wei, C., Li, Y., & Cai, X. (2011). Robust optimal policies of production and inventory with uncertain returns and demand. *International Journal of Production Economics*, 134(2), 357–367.
- Wen, R., Torkkola, K., Narayanaswamy, B., & Madeka, D. (2017). A multi-horizon quantile recurrent forecaster. In *31st conference on neural information processing systems* arXiv preprint arXiv:1711.11053.
- 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.
- Willemain, T. R., Smart, C. N., Shockor, J. H., & DeSautels, P. A. (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting*, 10(4), 529–538.
- Yanikoğlu, İ., Gorissen, B. L., & den Hertog, D. (2019). A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3), 799–813.
- Yelland, P. M. (2009). Bayesian forecasting for low-count time series using state-space models: An empirical evaluation for inventory management. *International Journal of Production Economics*, 118(1), 95–103.
- Zhao, X., Xie, J., & Leung, J. (2002). The impact of forecasting model selection on the value of information sharing in a supply chain. *European Journal of Operational Research*, 142(2), 321–344.
- Zheng, Y. S. (1992). On properties of stochastic inventory systems. *Management Science*, 38(1), 87–103.
- Zhu, S., Dekker, R., Van Jaarsveld, W., Renjie, R. W., & Koning, A. J. (2017). An improved method for forecasting spare parts demand using extreme value theory. *European Journal of Operational Research*, 261(1), 169–181.