ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Service-level anchoring in demand forecasting: The moderating impact of retail promotions and product perishability[☆]

Ben Fahimnia [a,*], Tarkan Tan [b], Nail Tahirov [c]

[a] Institute of Transport and Logistics Studies, The University of Sydney, Sydney, Australia
[b] Department of Business Administration, University of Zurich, Switzerland
[c] Department of Management and Marketing, Durham University Business School, Durham, UK

ARTICLE INFO

Keywords:
Demand forecast
Demand plan
Judgment
Service level
Promotion
Perishable
Laboratory experiment

ABSTRACT

The development of demand plans involves the integration of demand forecasts, service-level prerequisites, replenishment constraints, and revenue projections. However, empirical evidence has brought to light that forecasters often fail to distinguish between demand forecasts and demand plans. More specifically, forecasters frequently incorporate service-level requirements into their demand forecasts, even when explicitly instructed not to do so. This study endeavors to investigate the potential moderating impacts of product perishability and the presence of sales promotions on this phenomenon. Our findings reveal that sales promotions can meaningfully moderate the influence of service levels, since individuals tend to exhibit an elevated propensity for overforecasting during promotional periods when service levels are high. Surprisingly, no compelling evidence is found for the moderating effect of product perishability.

1. Introduction and background

In a demand-driven supply chain, a "demand plan" serves as the foundation for most tactical and strategic supply chain decisions. The demand plan encompasses (1) demand forecasts, (2) service-level requirements and replenishment constraints, and (3) revenue projections. Among these, demand forecasts hold particular significance. Demand forecasts are prepared by forecasters who gather information about the market (sales history and trends as well as contextual information such as the type/scale of sales promotions) to prepare their best sales prediction, which is a probability distribution that illustrates the possible sales scenarios (Goodwin & Fildes, 1999; Sanders & Ritzman, 1992). Subsequently, the demand forecast serves as an essential component for the demand planning process, which enables the release of the forecast to further planning and execution activities such as master planning, purchasing, allocation planning, and collaborative planning (Fahimnia, Pournader, Siemsen, Bendoly, & Wang, 2019). Importantly, forecasters rely primarily on historical data (i.e., previous sales, promotional data, and information related to special events and market dynamics) to develop demand forecasts, while demand planners consider a range of other factors (i.e., service-level requirements, product perishability, inventory obsolescence, and storage and shipping constraints) to prepare their demand plan.

The process of generating demand forecasts often employs statistical forecasting methods, primarily relying on

☆ The numerical results gathered in this manuscript were reproduced by the Editor-in-Chief on the 20th of October 2024.

* Corresponding author.
E-mail addresses: ben.fahimnia@sydney.edu.au (B. Fahimnia), tarkan.tan@business.uzh.ch (T. Tan), nail.tahirov@durham.ac.uk (N. Tahirov).

https://doi.org/10.1016/j.ijforecast.2024.07.007
0169-2070/

Please cite this article as: B. Fahimnia, T. Tan and N. Tahirov, Service-level anchoring in demand forecasting: The moderating impact of retail promotions and product perishability. International Journal of Forecasting (2024), https://doi.org/10.1016/j.ijforecast.2024.07.007.

historical data extrapolation. However, forecasters supplement these statistical tools with their expertise and intuition to incorporate the impact of contextual factors. These contextual factors relate to any event that could potentially prompt sudden fluctuations in demand, such as internal/external interruptions or competitors' actions (Abolghasemi, Hurley, Eshragh, & Fahimnia, 2020). This human intervention in the forecasting process is commonly referred to as "judgmental forecast adjustments". The judgment of the forecaster plays a crucial role in the forecasting process. Evidence shows that the rate at which forecasters adjust statistical forecasts at the SKU level can vary significantly across different contexts. While Franses and Legerstee (2009) reported adjustment rates around 90% in a pharmaceutical company, more recent studies suggest a broader range and provide more detailed insights. For instance, Fildes, Goodwin, and De Baets (2023) found that over 56% of forecasts were adjusted upwards, indicating a tendency towards positive adjustments depending on the dataset and industry. Many companies even employ a specific type of decision support system, namely, a forecasting support system (FSS), which combines a statistical forecasting approach with judgment from forecasters within the organization (Fildes, Goodwin, & Lawrence, 2006). In fact, managers may have access to information (e.g., potential promotional campaigns or competitor performance information) that is challenging to incorporate into a statistical model (Trapero, Pedregal, Fildes, & Kourentzes, 2013). Consequently, through judgmental adjustments, this type of information can be incorporated into the statistical model, which can enhance forecasting accuracy. Improvements in bias due to adjustments range from 37.5% to 77.5%, and improvements in forecast value added (FVA) range from 16.5% to 84.4% across different datasets (Fildes et al., 2023). These observations align with the findings of previous studies (Fildes & Goodwin, 2021; Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Syntetos, Nikolopoulos, & Boylan, 2010).

Our years of interactions with forecasters in the fast-moving consumer goods (FMCG) industry have revealed the pivotal role played by human judgment in forecasting, even when statistical forecasting tools are available. Survey results corroborate these observations, underscoring the importance of judgmental forecasting in various industry contexts (Diermann & Huchzermeier, 2017; Hofer, Eisl, & Mayr, 2015; Klassen & Flores, 2001; Sanders & Manrodt, 1994, 2003). Previous studies have reinforced the effectiveness of judgmentally made or judgmentally adjusted forecasts when individuals are adequately informed and/or trained (Alvarado-Valencia, Barrero, Önkal, & Dennerlein, 2017; Arvan, Fahimnia, Reisi, & Siemsen, 2018; Brau, Aloysius, & Siemsen, 2023; Harvey & Reimers, 2013; Ibrahim, Kim, & Tong, 2021; Seifert, Siemsen, Hadida, & Eisingerich, 2015).

In applying judgment, it is essential for forecasters to differentiate between useful and irrelevant information. One piece of information that has been shown to unduly influence the forecasters is target service levels. While demand forecasts should solely consider contextual information, such as promotions and special events, information related to service levels should be disregarded. Yet there is compelling evidence of a hidden anchor effect pertaining to service-level consideration in judgmental forecasting. Fahimnia, Arvan, Tan, and Siemsen (2023) demonstrated that demand forecasters evidently do account for service levels in their predictions, even when they are clearly instructed to focus on the most likely value of demand and incentivized to minimize the forecast error. In a series of laboratory experiments, the authors established that this anchoring effect is driven particularly by service-level information, and not some other anchor.

Another critical factor in forecasting is the influence of promotional activities, which add complexity by challenging forecasters to manage many variables, often with limited data. This complexity underscores the necessity of discerning essential information, especially when balancing service-level objectives with promotional activities (Fildes, Ma, & Kolassa, 2022). We contend that the consideration of service-level targets may be even more pronounced during promotional periods. Promotions are not always profitable, and in fact evidence from the FMCG industry indicates a significant increase in stockout rates during promotional periods (ECR Australia, 2010). Consequently, demand planners understandably take service-level information more seriously during promotional periods. Forecasters often ignore service-level targets, as their main focus is on the accuracy of forecasts rather than maintaining these levels (Oliva & Watson, 2011). This tendency is even more pronounced during promotional periods when the benefits of promotional information for forecasting accuracy (Lei et al., 2023; Ma & Fildes, 2021; Ma, Fildes, & Huang, 2016) lead them to further overlook service-level targets. Our extensive experience collaborating with forecasters in the FMCG industry does not align with this perspective. We argue that the service-level anchor reported by Fahimnia et al. (2023) would be even more pronounced during promotional periods.

Accurate forecasting is especially crucial for businesses selling products with limited shelf lives, as inaccuracies can lead to environmental damage and financial losses from disposing of unsold goods (Eiglsperger et al., 2024; Huber, Gossmann, & Stuckenschmidt, 2017). FMCGs include perishable items that deteriorate quickly or have an expiration date. Maintaining a high service level for these products often results in excessive inventory leftovers when sales are overestimated (Van Donselaar, Peters, Jong, & Broekmeulen, 2016). In this respect, we also argue that product perishability may play a moderating role in factoring service-level information in demand forecasting. Forecasters tend to overforecast when informed about high service levels, yet overforecasting perishable products prompts waste and sales markdowns. Given that waste considerations are paramount in inventory decisions, forecasters may be similarly influenced by product perishability information when making judgmental forecasts (Sanders & Ritzman, 2001a). There is a dearth of evidence in the existing literature on whether product perishability can impact the overforecasting behavior induced by high service-level targets.

Our aim in this paper is to contribute to this literature by investigating whether the established impact of service-level targets on demand forecasts (Fahimnia et al., 2023) is moderated by sales promotions and product perishability information. The primary questions we wish to address in this paper are as follows: (1) To what extent does the presence of *sales promotions* moderate the impact of service-level considerations in demand forecasting? (2) To what extent does *product perishability* moderate the impact of service-level considerations in demand forecasting?

To address these questions, we designed a laboratory experiment that mimics the demand forecasting process in two large FMCG companies in Australia. Empirical data related to sales history and promotional events were gathered from these companies. The data were manipulated to produce a unique time series (statistically replicable) for each subject. We collected data from 313 subjects who were fresh graduates (less than 15%) and experienced forecasters (over 85%) in Australia and the Netherlands.

2. Literature and hypothesis development

The target service level acts as a crucial bridge between forecasts and order quantities. Particularly when facing uncertain demand, maintaining an adequate service level is a primary concern for managers and planners. To achieve this, managers can determine and implement an optimal inventory strategy, such as holding safety stock, supported by accurate demand forecasts. Accurate forecasting is especially critical for addressing newsboy problems, which involve determining the optimal inventory level for perishable products with uncertain demand in a single production or procurement setting. Managers can reduce the risk of overstocking or inventory shortages by developing forecasting models tailored to specific problems (Artto & Pylkkänen, 1999) or by implementing schemes for updating forecasts (Lee, 2008; Tiwari, Patil, & Shah, 2011). These approaches help in meeting the target customer service level (Syntetos, Boylan, & Disney, 2009). Inventory and order managers, therefore, combine demand forecasts with the target service levels to determine optimal order quantities. In practice, this process often requires addressing various challenges and complexities.

In demand-driven supply chains, sales and operations planning (S&OP) serves as the central organizational mechanism for integrated planning (tactical and strategic supply-chain decision making), integrating decision making with demand forecasts. In the S&OP literature, accurate forecasting is identified as a crucial component for facilitating integrated planning (Ivert & Jonsson, 2010; Kaipia, Holmström, Småros, & Rajala, 2017; Nakano, 2009; Oliva & Watson, 2011). S&OP teams deal with developing an aggregate plan aimed at meeting forecasted demand by adjusting production rates, inventory levels, labor levels, and other controllable variables, in accordance with the target customer service level (DuHadway & Dreyfus, 2017; Heizer, Render, & Munson, 2017). Though forecasting and planning are both future-oriented decision-making processes, forecasting focuses on accuracy (Alvarado-Valencia & Barrero, 2014), while planning takes a more comprehensive approach, considering various factors, constraints, and limitations to ensure effectiveness (Hogarth & Makridakis, 1981). Service-level consideration falls within the realm of planning rather than forecasting (Sodero, 2022). In other words, service levels should be incorporated during the planning stage within S&OP, leaving the forecasting stage unaffected by service-level information.

Human intervention in forecasting (known as judgmental forecasting or judgmental forecast adjustment) comes with personal biases (Harvey & Fischer, 1997; Önkal, Goodwin, Thomson, Gönül, & Pollock, 2009). Anchoring (Tversky & Kahneman, 1974), wishful thinking (Morlock, 1967), illusions of control (Langer, 1975), and hindsight bias (Fischhoff, 1975) are examples of these biases. Judgmental adjustments are typically employed when contextual information, such as expected reactions from competitors, is challenging to incorporate into the statistical forecast (Perera, Hurley, Fahimnia, & Reisi, 2019). Research confirms the effectiveness of human intervention in forecasting (Alvarado-Valencia & Barrero, 2014; Alvarado-Valencia et al., 2017; Arvan et al., 2018; Broeke, Baets, Vereecke, & Baecke, 2019; De Baets & Harvey, 2020a, 2020b, 2020c, 2020d; Fildes & Goodwin, 2007, 2021; Harvey & Reimers, 2013; McCarthy, Davis, Golicic, & Mentzer, 2006; Perera et al., 2019; Seifert et al., 2015; Syntetos et al., 2010).

For example, in their survey of 149 forecasters, Fildes and Goodwin (2007) demonstrated that only 25% of forecasts rely exclusively on statistical methods, and that it is common for forecasters to make subjective adjustments to quantitative forecasts. Furthermore, their results highlight the significance of judgment in forecasting, evidenced by the majority of respondents acknowledging its importance, and 34% emphasizing that judgment is crucial. A synthesis of existing surveys summarized by Sroginis, Fildes, and Kourentzes (2023) showed that around 40% of responses indicate the use of judgmentally adjusted statistical forecasts, confirming the ongoing importance of combining statistical methods with contextual insights. Alvarado-Valencia and Barrero (2014) conceptualized research on judgmental forecasting and concluded that the use of heuristics and the reliance on computer-generated suggestions are key human behaviors influencing the forecasting task. Broeke et al. (2019) studied the variability of judgmental forecast adjustments across different time horizons and their influence on forecast accuracy. They discovered that adjustments made closer to the sales point tend to increase in size and positivity, but these changes do not consistently enhance forecast accuracy. Their findings highlight the context-specific nature of the effectiveness of such adjustments. Fildes, Goodwin, Önkal, and Thomson (2009) conducted an empirical investigation into the effects of judgmental adjustments on computer-generated forecasts within supply chain planning across four companies. They analyzed over 60,000 forecasts and found that larger adjustments generally lead to improvements in accuracy, while smaller adjustments usually result in diminishedaccuracy. Similarly, Syntetos et al. (2010) investigated the effects of judgmental adjustments on statistical forecasts, focusing on their impact within inventory forecasting. Their study, which utilized data from the pharmaceutical industry, demonstrated that such adjustments not only improve forecast accuracy but also significantly influence stock levels and service levels.

The primary objective in demand forecasting revolves around minimizing error, and as such, all units of forecast error should appear the same to a forecaster. However, it has been shown that forecasters inadvertently use loss functions to compare the cost of underforecasting with the cost of overforecasting, particularly when influenced by service-level information. Empirical studies by Fildes and Goodwin (2007) revealed that the majority of forecasters (63.9%) consider underforecasting to be more costly. Separately, Fildes (2006) found that forecasters overforecast demand by an average of 24.7%.

Fahimnia et al. (2023) demonstrated that service levels act as implicit anchors in demand forecasting. They are not explicit anchors because service levels are expressed as percentages whereas forecasts tend to be in units of product. They act as implicit anchors because they may lead forecasters to consider a portion of the probability distribution different from the mean or the median; point forecasts may thus not represent the median of a distribution but may become upward- or downward-biased, depending on the target service level. A service-level target[1] above 50% anchors decision makers in an area of the underlying demand distribution higher than the median. Service levels in the FMCG context are well above 50%. In our experiment, we considered a Type I service-level target of 98.5% as a "high service-level target".[2] The lowest service level we have observed in this industry is 85%; therefore, we refer to a service-level target of 85% as a "lower service-level target" or "service-level target that is perceived as low" in the FMCG industry. This definition and distribution between high and low service levels is essential in our study given that we used experienced forecasters and supply chain practitioners as participants in our experiments. This is not how high and low service levels were defined by Fahimnia et al. (2023). With this background, we hypothesize:

H1 A. A high service-level target creates overforecasting bias.

H1B. The higher the service-level target, the more pronounced the overforecasting bias.

Service-level information has the potential to play a more significant role during promotional periods. The impact of promotions has also been extensively studied in the forecasting literature (De Baets & Harvey, 2018; Fildes, Goodwin, Önkal, & Lawrence, 2019; Hewage, Perera, & De Baets, 2022; Ma et al., 2016; Sroginis et al., 2023; Trapero, Kourentzes, & Fildes, 2015). Trapero et al. (2015), for instance, proposed a forecasting model utilizing principal component analysisto efficiently handle promotional challenges in demand forecasting. This model outperformed traditional expert and statistical forecasts by accurately predicting sales for products with limited history through data pooling. Trapero et al. (2013) examined the accuracy of managerial adjustments to forecast during promotions, using data from a manufacturing company. They observed that while judgmental adjustments have the potential to enhance forecast accuracy during promotions, the magnitude of these adjustments is frequently too large, negatively affecting the overall accuracy. De Baets and Harvey (2020a) investigated the effectiveness of various forecasting support strategies in the context of time series affected by sporadic promotions. In contrast to Trapero et al. (2013), they found that underforecasting during promotions and overforecasting in normal periods are prevalent, regardless of the forecasting method. Notably, the study also found that providing forecasters with optimal statistical forecasts which fully incorporate promotional effects significantly enhances accuracy by almost eliminating biases and reducing random error by 20%.

Sroginis et al. (2023) explored the role of human judgment in adjusting promotional forecasts, focusing on the utilization of contextual and model-based information. Their findings show that forecasters often rely on model-based benchmarks, such as recent promotional uplifts and current statistical forecasts, sometimes ignoring important contextual information about past promotions. The study further revealed that when working with promotional forecasting models, forecasters often misinterpret information, leading to excessive adjustments and reduced accuracy. Fildes et al. (2019) studied the challenges of forecasting sales promotions within supply chains, emphasizing the complexity of adjusting baseline forecasts with judgmental inputs. The study found that giving forecasters promotional information with unclear value can harm forecast accuracy, indicating a need to redesign forecasting systems to only include clear, relevant information to enhance accuracy.

Promotions are an important part of the marketing mix in the FMCG industry (Hewage et al., 2022), as a considerable portion of sales occur during cyclic promotions. Even for medium-size enterprises, the costs associated with unmet service levels during promotions can amount to millions of dollars (Craig, Dehoratius, & Raman, 2013). Nonetheless, promotions are not always profitable, with only about 18% of promotions falling into that category (Srinivasan, Pauwels, Hanssens, & Dekimpe, 2004). A survey of the FMCG industry carried out by Efficient Consumer Response Australasia (ECRA) reported that stockout rates significantly surge in promotional periods, making promotions less profitable than they are perceived to be (ECR Australia, 2010). Although maintaining service levels

[1] Note that we focus on Type I service levels in our research because this is a commonly applied concept in the context of FMCG.

[2] This is the service-level requirement used by the retailers from whom we collected the data for our experiments.

in promotions is a logical consideration in demand planning and S&OP, a forecaster who is repeatedly exposed to this information may naturally place greater emphasis on target service levels during promotional periods (while service-level information should be entirely ignored regardless of promotions). We develop our next hypothesis on this basis:

H2. Service-level targets influence forecasts more strongly during promotional periods than regular periods.

Forecasting for perishable products, particularly in the food industry, introduces additional complexities (Perera et al., 2019). Studies focusing on product perishability in the field of forecasting research are relatively rare (Fildes et al., 2022). Van Donselaar et al. (2016) examined the effect of relative price discounts on the sales of perishable products during promotional periods and found that consumer responses to promotions vary significantly, due to differing shelf lives. Khosrowabadi, Hoberg, and Imdahl (2022) discovered that adjustments to AI-generated demand forecasts are common for perishable products, especially during promotions or unusual weather conditions. While large positive adjustments often lack accuracy, large negative adjustments generally improve it. Surprisingly, planners do not, on average, enhance forecast accuracy through their adjustments. Underforecasting risks lost sales and reputational damage, while overforecasting prompts waste and sales markdowns. Food manufacturers and retailers have grappled with the consequences, as food waste remains a major concern in supply chains, with an estimated 50% of all food produced being wasted before and after reaching the consumer (Lundqvist, De Fraiture, & Molden, 2008). Food chain experts attribute a substantial portion of this waste to forecasting and planning inadequacies (Mena, Adenso-Diaz, & Yurt, 2011), a hot discussion topic in the media and in executive publications/seminars/forums. Waste consideration is imperative in the classic newsvendor problem (Petruzzi, 1999) where optimal inventory decisions depend on waste generation, lost sales, holding costs, and capacity constraints (Qin, Wang, Vakharia, Chen, & Seref, 2011).

Given the significant implications of perishability for organizations, especially food producers, forecasters are likely influenced by product perishability when making judgmental forecasts. This influence can counteract the overforecasting behavior induced by high service-level targets (see Sanders & Ritzman, 2001b). Therefore, we hypothesize:

H3. High service-level targets lead to less overforecasting for perishable products.

Section 3 presents the design of our laboratory experiment, which will be employed in Section 4 to collect data and test these hypotheses.

3. Experimental design

One major strength of experimental research is its ability to facilitate observations in simplified, controlled conditions, as opposed to the complexity of natural settings. Despite criticisms about its limited ability to generalize to real-world contexts (i.e., having external validity; Thye (2014)) or to reveal all aspects of the behavior (Franses, 2013), empirical studies (e.g., Dipboye & Flanagan, 1979; Locke, 1986) have shown that experimental results can indeed be broadly applicable. Controlled laboratory experiments are the most common research approach in the literature on judgmental forecasting (Arvan et al., 2018; Bendoly, Donohue, & Schultz, 2006). An important aspect of our experiment is that we used a mix of students (15%) and experienced forecasters (85%) from the FMCG industry as participants.

The experiment was devised in partnership with practitioners who have daily engagement in forecasting tasks. They reviewed and refined the task description, time series formatting, and other aspects through multiple iterations before the actual experimental began. During our preliminary tests (prior to each experiment), we explicitly informed the participants that they would be engaging in a demand forecasting task. We emphasized the need to disregard all external factors typically considered in demand planning tasks. These test runs provided an opportunity for participants to pose questions and gain clarity on the task at hand before the commencement of the actual experiment.

The experiment began by providing all subjects with general information about the case industry and their role as forecasters. The difference between demand forecasting and demand planning and supply-chain decision making was explicitly reinforced in the task description (Fig. A.1 in Appendix A illustrates an example of the task description). More precisely, all subjects were adequately notified—through written and verbal instructions—that their forecasts should reflect only the most likely value for product demand in the forthcoming week based on the historical sales data and possible sales promotions, and that their rewards will depend on the accuracy of their forecasts. They were also told that their forecasts would then be forwarded to other departments where additional factors—including internal/external constraints and requirements—would be taken into consideration to make related supply chain decisions.

Each subject was assigned to one of six treatment groups (Fig. 1). In Treatment 1 (T1) and Treatment 2 (T2), no service level information is revealed to the subjects, and the forecasts were made for a nonperishable product (shelf life of 9 months in T1) and a perishable product (shelf life of 1 day in T2). Treatment 3 (T3) aimed to test the impact of high service level (testing *H1A*). In this treatment, forecasts were made for a nonperishable product, and the subjects were informed about a high service level of 98.5%. In Treatment 4 (T4), the forecasts were still for a nonperishable product, but the subjects were informed of a lower service level requirement of 85% (testing *H1B*). Finally, Treatments 5 (T5) and 5 (T6) were characterized by forecasting for a highly perishable product (testing *H3*) in which the service level was either high (T5) or low (T6).

A total of 368 subjects prepared four forecasts each (Table 1). For each forecast, a subject was provided with 30 weeks of sales data with both normal and promotional weeks. The promotional weeks were highlighted as "Promo". The subjects were asked to provide their

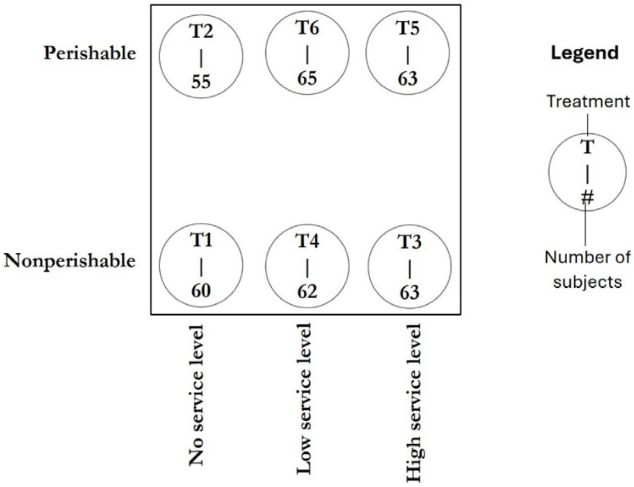


Fig. 1. Overview of treatment design.

Table 1
Initial statistics of the number of subjects and their overall performance.

Treatment	No. of subjects		Percentage forecast bias		MAPE	
			Mean	Median	Mean	Median
Treatment 1 (no SL, nonperishable)	60	Nonpromotional period	−0.3%	−0.6%	5.45%	4.60%
		Promotional period	−1.2%	−0.2%	3.52%	2.95%
Treatment 2 (no SL, perishable)	55	Nonpromotional period	−0.7%	−0.7%	4.47%	3.47%
		Promotional period	−4.6%	−5.9%	7.41%	7.24%
Treatment 3 (high SL, nonperishable)	63	Nonpromotional period	+3.4%	+1.9%	6.21%	4.29%
		Promotional period	+7.2%	+7.2%	7.72%	7.35%
Treatment 4 (lower SL, nonperishable)	62	Nonpromotional period	−3.9%	−4.7%	6.44%	5.88%
		Promotional period	−2.9%	−2.9%	4.20%	3.75%
Treatment 5 (high SL, perishable)	63	Nonpromotional period	+5.1%	+4.7%	5.69%	5.02%
		Promotional period	+5.9%	+5.7%	7.64%	6.51%
Treatment 6 (lower SL, perishable)	65	Nonpromotional period	+2.0%	+1.4%	5.27%	4.41%
		Promotional period	−5.4%	−7.0%	9.93%	9.36%

forecasts for week 31 and instructed to base their forecast solely on historical data and potential sales promotions. In two out of four attempts, the subjects forecast for a promotional period (testing *H2*). That is, for half of their forecasting tasks, participants were specifically predicting sales outcomes for a week marked by promotional activities. Subjects were told they could assume that the impacts of promotions on sales are independent from each other. Additional information, tailored specifically to each treatment group, about product shelf life and retailers' service levels was provided to the forecasters.

The experiment started with a set of questions to collect demographic information from the subjects related to their gender (male: 60%; female: 40%), age (18–25 years: 22%; 26–30 years: 34%; 31–35 years: 33%; 36–45 years: 10%; 46–55 years: 1%), qualifications (postgraduate: 49%; bachelor/honors: 43%; other: 8%), forecasting experience (1–2 years: 28%; 2–5 years: 28%; 5–10 years: 7%; less than 1 year: 18%; more than 10 years: 5%; no experience: 14%), and related work experience (1–2 years: 25%; 2–5 years: 14%; 5–10 years: 7%; less than 1 year: 18%; more than 10 years: 13%; no experience: 23%). At the end of the experiment, the subjects were asked to state the factors they considered when making the forecasts. The options consisted of (1) historical sales data, (2) past promotional information, (3) upcoming promotions, (4) product shelf life, (5) retailer service level, (6) seasonality in the past sales, (7) trends in the past sales, (8) noise (sudden fluctuations) in the past sales, and (9) personal industry insights.

For each forecast, a subject received a unique historical sales dataset. We used real data from two huge food and beverage companies to generate the historical sales data. The sales data and promotional information for the nonperishable product (nine-month shelf life) were obtained from a large beverage company. Data for the perishable product (one-day shelf life) were obtained from a large bread manufacturing company. The characteristics of the real datasets were replicated in all sales data. Such parameters as noise, frequency of promotions, and the impact of promotions on sales uplift were estimated to replicate the real data characteristics. Therefore, each subject received four unique datasets to produce four forecasts (two promotional and two nonpromotional weeks). Below are the equations used to produce the required datasets (Kremer, Moritz, & Siemsen, 2011):

$$S_t = \mu_t + \epsilon + x_t(\beta + \delta) \tag{1}$$

$$\mu_t = \mu_{t-1} \tag{2}$$

$$S.t.: \sum_{t}^{N} x_t = f \tag{3}$$

In Eq. (1), S_t and x_t represent sales in week t and the binary promotion variable in week t, respectively. ϵ and δ are normally distributed independent random variables with zero mean, resembling random noise affecting demand and promotions, respectively. The promotional impact is shown by β. Eq. (2) calculates the seed for generating random sales figures. The constraint ensures that only f promotions happen in the entire forecasting horizon, where f has a Poisson distribution. All parameters in Eqs. (1), (2), and (3) were estimated from the real sales data.

The normative benchmark forecast is used as a baseline to evaluate the effectiveness of judgmentally adjusted forecasts. Comparing the adjusted forecasts to this benchmark allows us to assess whether the adjustments improve forecast accuracy. To encourage their efforts, we gave monetary incentives to all subjects. All participants received a show-up fee of \$5 as well as an additional payment of up to \$10 depending on the accuracy of their forecasts (total payments between \$5 and \$15). The mean absolute percentage error (MAPE) was used to assess the accuracy of the forecasts (Broeke et al., 2019; Fildes & Goodwin, 2007; Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; McCarthy et al., 2006). The MAPE was calculated using Eq. (4), where F_t^N and F_t are the normative benchmark forecast and the judgmentally adjusted forecast, respectively, for period t, and n is the total number of forecasts:

$$MAPE = 100\frac{\sum_t \left|\frac{F_t^N - F_t}{F_t^N}\right|}{n} \tag{4}$$

The normative benchmark forecast (F_t^N) that we used is the average of historical data from previous nonpromotional periods (i.e., the series presented to each subject). For the forecasts in promotional periods, we adjusted this benchmark by adding the population mean of the uplift factor used, in order to reduce noise in the outcome metric (see Fahimnia et al., 2023). Calculating forecasting performance using this approach in behavioral experiments with stationary data is common (e.g., Cerqueira, Torgo, & Mozetič, 2020; Lee & Siemsen, 2016). Our analyses and results remained consistent regardless of whether errors were calculated as deviations from actuals or deviations from the normative benchmark. This approach ensures that our analysis adheres to established methods and maintains consistency.

4. Analysis

4.1. Data collection and initial statistics

The data for testing our hypotheses were collected in 15 experimental sessions consisting of three sessions for each of the five treatments. As stated above, each subject made four forecasts (two for a promotional period, and two for a nonpromotional period). To ensure the robustness of our statistical analysis, we systematically identified and excluded outliers from our dataset. Outliers were defined based on the interquartile range (IQR) method. Specifically, data points below the first quartile (Q1) minus 1.5 times the IQR or above the third quartile (Q3) plus 1.5 times the IQR were considered outliers and removed. Table 1 presents the number of subjects and their overall performance. For example, 60 subjects completed Treatment 1, generating 240 data points (120 each for nonpromotional and promotional periods). Additionally, a more detailed analysis that disaggregates performance metrics such as the MAPE and percentage forecast bias for individual subjects can be found in Appendix B, under Fig. B.1.

Statistical forecast bias is defined as the degree, on average, by which a forecast deviates from the actual sales point. Various metrics have been introduced to statistically measure the forecast bias. We adopted the metric introduced by Petropoulos, Fildes, and Goodwin (2016), which calculates the relative deviation from the normative benchmark forecast (F_t^N) detailed in Section 3. The

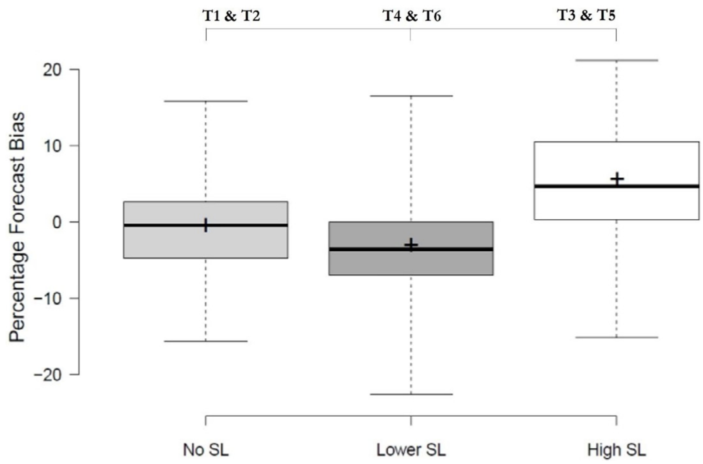


Fig. 2. Percentage forecast bias at various service levels.

percentage forecast bias was calculated using Eq. (5), under the assumption that $F_t^N > 0$. This measure is scale-free and easy to interpret. A percentage bias of 0% means there is no deviation from the normative benchmark forecast. Negative and positive numbers imply underforecasting and overforecasting, respectively.

$$\text{Percentage forecast bias} = 100(\frac{F_t}{F_t^N} - 1)\% \tag{5}$$

We depict in Fig. 2 the percentage forecast bias at various service levels in the presence and absence of promotions. Service-level requirements significantly affect the forecast bias (when other factors are averaged out); see Fahimnia et al. (2023) for empirical support. This effect is significant, as supported by the ANOVA test presented in Table 2. Furthermore, promotions are also shown to have a significant influence on forecast bias. Additionally, a separate ANOVA test analyzing forecast accuracy is shown in Table 3, further underscoring the influence of promotions on forecast outcomes. The interaction effect between the service level and promotions is also significant in both analyses. Although perishability information alone does not lead to a significant impact on forecast bias, the interactions between "service level and perishability" and "promotions and perishability" do have significant impact on both forecast bias and accuracy. Note that to ensure the independence of observations, we averaged the two promotional and nonpromotional forecasts for each subject, which resulted in two independent observations per subject. Degrees of freedom (df_1: between-subject degrees of freedom; df_2: within-subject degrees of freedom) for the repeated-measures ANOVA are indicated in Tables 2 and 3.

4.2. Hypothesis testing

The following section presents the results of our hypothesis testing, specifically analyzing the impact on the forecast bias using t-tests. The overall findings, including t-ratios and p-values from these tests, are summarized in Table 4. The table outlines how service levels, sales promotions, and product perishability influence forecast bias, detailing the effects and interactions among these key factors.

Comparing the forecast bias in high service level treatments with no service level treatments (i.e., T3 vs. T1, and T5 vs. T2), we found that revealing a high service level requirement to the subjects resulted in a highly significant forecast bias: A high service level requirement results in overforecasting. Accordingly, H1A is supported.

Comparing the forecast bias in treatments with lower service levels to no service level treatments (i.e., T4 vs. T1 and T6 vs. T2) indicates that a service level of 85% has an insignificant impact on forecast bias. Because most retailers perceive a service level of 85% as (very) low, we found it no surprise that it triggers no overforecasting. Nevertheless, this result supports H1B because only the higher service level is associated with overforecasting bias.

The forecast bias comparison between forecasting for promotional and nonpromotional weeks in T3 and T4 shows that the presence of sales promotions moderates the effect of service level information on forecast bias: When the service level is high (T3), the subjects are more biased toward overforecasting when forecasting for promotional weeks. In contrast, when the service level is low (T4), there is more of a tendency to underforecasting in the promotional weeks, as plotted in Fig. 3. Accordingly, H2 is supported.

The results do not confirm the moderating impact of product perishability on service-level consideration. The significant forecast bias caused by high service-level requirements remained statistically unchanged when the product shelf life was changed from nine months to one day, as plotted in Fig. 4. When the product shelf life was changed in the low service-level scenario, the changes in the percentage forecast bias were not meaningful either. Accordingly, *H3* is not supported.

Our analysis of the demographic data collected from all subjects before the experiment found no connection between qualifications (i.e., last obtained degree) and average forecast bias, but there was a slight moderation

Table 2
ANOVA test results for percentage forecast bias (over- vs. underforecasting).

Factor	df_1	df_2	F	p-value
Service level	2	355	97.03	<.0001***
Perishability	1	355	0.00	0.99
Promotions	2	355	10.56	0.001**
Service level: Perishability	1	355	3.16	0.04*
Service level: Promotions	2	355	20.55	<.0.0001***
Perishability: Promotions	1	355	39.14	<.0.0001***
Service level: Perishability: Promotions	2	355	3.94	0.02

*p ≤ 0.05; ** p ≤ 0.01, ***p ≤ 0.001.

Table 3
ANOVA test results for forecast accuracy (MAPE).

Factor	df_1	df_2	F	p-value
Service level	2	355	8.77	0.0002***
Perishability	1	355	12.57	0.0004***
Promotions	2	355	12.38	0.0005***
Service level: Perishability	1	355	5.49	0.004**
Service level: Promotions	2	355	1.10	0.33
Perishability: Promotions	1	355	47.11	<.0001***
Service level: Perishability: Promotions	2	355	11.03	<.0001***

*p ≤ 0.05; ** p ≤ 0.01, ***p ≤ 0.001.

Table 4
Outcomes of statistical tests to examine the developed hypotheses.

Hypothesis		t-ratio	p-value
H1A Overforecasting	Lower service level	-1.258	0.4200
	High service level	11.293	<.0001
H1B Difference between service levels	Lower service level	12.786	<.0001
H2: Moderation impact of sales promotions	Lower service level	5.259	<.0001
	High service level	-3.275	0.0012
H3: Moderation impact of product perishability	Lower service level	-1.694	0.0911
	High service level	-0.356	0.7219

impact on service-level consideration (F = 2.72, *p*-value = 0.07). Nevertheless, we noticed a significant difference between undergraduate and postgraduate degree holders in consideration of a high service level: Subjects with undergraduate degrees did not overforecast as much as those with postgraduate degrees (t-ratio = −2.59, *p*-value = 0.01). This is supported by research in the existing literature showing that formal education has a negative effect on forecasting performance, or is, at best, a non-contributor (Šindelář, 2016). Related industry experience also has a moderating impact on service-level consideration. Junior forecasters (1–5 years of industry experience) are more biased in demand forecasting than inexperienced subjects. However, we also noted that subjects with more substantial industry experience (5+ years) performed comparably to fresh graduates (t-ratio = −1.773, *p*-value = 0.3910).

4.3. Discussion

In this paper, we extended the study of Fahimnia et al. (2023) to further explore the extent to which demand forecasters take service-level requirements into account when developing their base forecasts. Our findings reaffirm the anchoring effect of service-level information

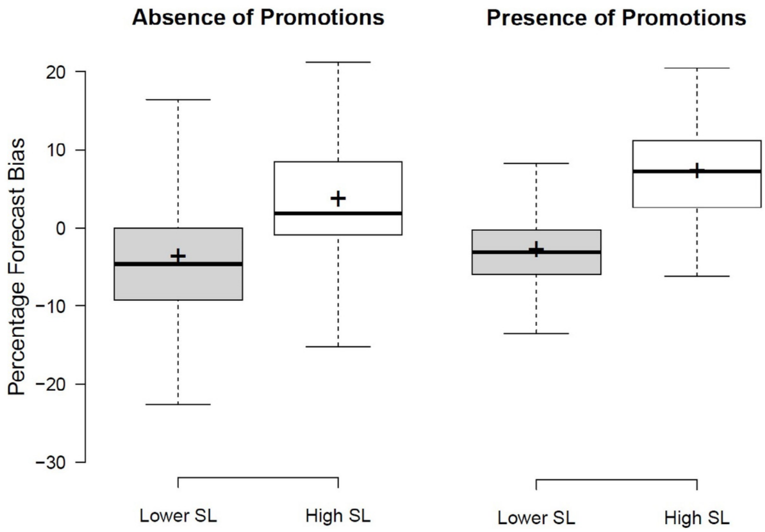


Fig. 3. Impact of promotions on percentage forecast bias at various service levels.

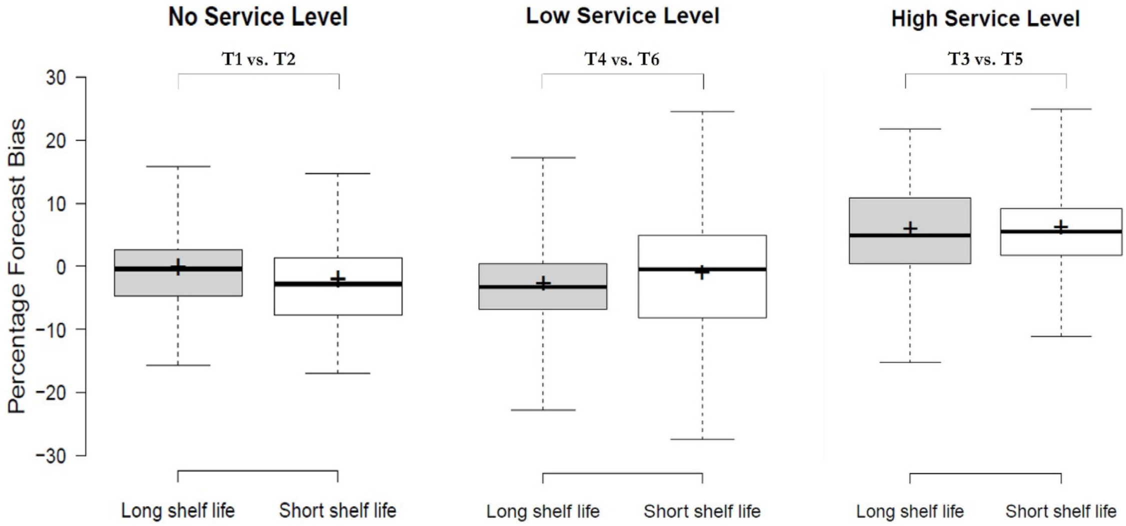


Fig. 4. Impact of product perishability on percentage forecast bias at various service levels.

initially identified by Fahimnia et al. (2023). Despite explicit instructions, the forecasters in our experiments evidently factored in service-level information when preparing their demand forecasts. The percentage forecast bias when a high service level was required turned out to be 9% higher than that when a required service level is perceived as low (105.4% for a high service level vs. 96.6% for a low service level).

We further investigated whether sales promotions and product perishability information moderate the impact of service-level information. It turned out that in promotional periods, when the service level is high, the subjects display a stronger bias toward overforecasting, but when the service level is lower, the tendency is toward underforecasting. This behavior may be attributed to how the service level is perceived as an indicator of a retailer's performance (a high service level implying a better-performing retailer). The subjects did not seem to trust a retailer with lower service-level targets to run successful promotional campaigns, and accordingly tended to underforecast the demand. Conversely, they raised their forecasts when dealing with retailers perceived as having higher service levels, expecting successful promotions with them. Accordingly, in promotional periods, the percentage forecast bias exhibited a difference of over 9.4% under varying service-level requirements (106.5% during promotional periods vs. 97.1% during nonpromotional periods).

However, we did not find evidence for the moderating impact of product perishability on service-level consideration. This is contrary to common expectations, particularly in light of the growing global concerns about food waste and the anticipated influence on human judgment. Forecasters tend to rightfully disregard perishability information in forecasting, while service-level anchoring persists (H1A/H1B in our study; and Fahimnia et al., 2023) and the impact is even further amplified during promotional periods (*H2*). One plausible hypothesis is that the correlation between a short shelf life and waste generation may not be immediately evident, or the economic/environmental implications may be unclear to the forecasters. Empirical testing of such hypotheses could be a promising avenue for future research in this area.

The qualification and experience of the subjects did not consistently produce significant effects on the results. Prior studies have found no significant difference between using student subjects and practitioner subjects (Bolton, Ockenfels, & Thonemann, 2012). For instance, Fahimnia et al. (2023) incorporated both students and practitioners in their experiments and found no significant variation in outcomes. We used practitioners to ensure that our subjects possessed the relevant experience and foundational knowledge in the field. While qualifications had a negative impact on overforecasting under high service-level requirements (resulting in higher forecast bias), industry experience had a positive effect. Interestingly, subjects with substantial industry experience (5+ years) demonstrated a proficiency on par with fresh graduates in distinguishing between demand forecasting and demand planning. Overall, we believe that our findings are robust to the subjects' demographics.

5. Conclusions

The demand plan is a critical input in the sales and operations planning (S&OP) process. Demand planners utilize demand forecasts developed by forecasters, along with factors such as service levels, stock levels, and supply/shipping constraints, to prepare their demand plans. This collaborative process is particularly essential under uncertain conditions like new product launches or promotions. However, in real-world business environments, demand forecasters face significant challenges in developing accurate forecasts, due to a lack of information, unexpected market events, or deviations from their primary tasks (DuHadway & Dreyfus, 2017). For instance, traditional point-of-sale (PoS) information sharing can be costly and is typically only useful in specific settings where traditional demand planning is inadequate, such as during product launches, promotions, and seasonal peaks (Kaipia et al., 2017). Therefore, developing new data-sharing procedures tailored to specific industries can significantly enhance forecasters' ability to achieve accurate forecasting. In the S&OP process, it is crucial to match supply and demand by incorporating demand information from forecasters who consider marketing dynamics like promotions. This ensures integrated operations planning and significantly impacts a firm's performance (Sodero, 2022).

The demand forecast, as the fundamental input into the demand plan, should not take a loss function into account because its sole purpose is to provide the best possible estimate of sales. Demand forecasting requires sales history, but these data alone are insufficient due to the multitude of internal and external variables impacting sales, including special events and promotional strategies. Therefore, integrating human intervention into forecasts seems inevitable, as such factors are not easily quantifiable in statistical forecasting models.

This study can be extended in various ways, presenting an opportunity for future research. Our study focused on exploring the presence and attributes of judgmental forecast adjustments, rather than their impact on forecast accuracy when combined with statistical forecasts, a topic extensively studied in the literature (e.g., Broeke et al., 2019; Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Syntetos et al., 2010). Judgmental forecasting is subjective by definition, with inherent limitations such as inconsistency and bias. Research has shown that judgment can be unreliable, leading to inconsistency (i.e., unsystematic deviations from the optimal forecast; Stewart (2001)). Strategies such as limiting the use of judgment in forecast adjustments have been used to mitigate these issues (Sanders & Ritzman, 2001a; Webby, O'Connor, & LaWrence, 2001). However, empirical evidence suggests that managers frequently and significantly deviate from statistical forecasts, without necessarily improving forecast accuracy, due to factors such as the diverse and dynamic nature of loss functions (Franses, 2013; Petropoulos et al., 2016), data characteristics and complexity (Baecke, De Baets, & Vanderheyden, 2017; Lawrance, Goodwin, O'Connor, & Önkal, 2006), the expertise effect (Mary E. Thomson & Macaulay, 2004), special events (Goodwin & Wright, 2010; Nikolopoulos, Litsa, Petropoulos, Bougioukos, & Khammash, 2015), individual differences (Eroglu & Croxton, 2010; Eroglu & Sanders, 2020), and lack of support from top managers (Fildes & Goodwin, 2007; Goodwin, 2000). Employing a systematic and well-structured approach in judgmental forecasting can help mitigate such limitations. Strategies such as employing checklists of relevant information categories, maintaining records of forecasts (Harvey, 2001), defining rational loss functions, and engaging in group forecasting (i.e., as group forecasting can leverage collective expertise and diverse perspectives, potentially reducing individual biases and inconsistencies) can be utilized (Lawrance et al., 2006; Lawrence, O'Connor, & Edmundson, 2000). There are, however, contradictory findings regarding the presentation of data. Some studies have suggested that providing data in graphical form may better estimate trends, while others advocate for tabular formats to reduce bias (Desanctis & Jarvenpaa, 1989; Dickson, DeSanctis, & McBride, 1986; Harvey, 2001; Harvey & Bolger, 1996; Lawrance et al., 2006; Lawrence, 1983)). Trends are often better estimated from graphical presentations, but these can also encourage inconsistency and overforecasting compared to tabular formats. In light of these findings, future research could extend our work by considering more realistic loss functions, by exploring various modes of task presentation (i.e., tabular and graphical formats), and by incorporating group forecasting methods into experimental designs.

Given our findings that product perishability does not play a moderating role in demand forecasts, future research should examine how forecasters perceive and integrate perishability into their decision-making processes, considering external factors like weather, promotion, and competitive market dynamics (Chen & Ou, 2009). Furthermore, experimental studies could investigate the impact of providing forecasters with varying degrees of perishability information, including differences in shelf life (Van Donselaar et al., 2016), consumer freshness preferences, and environmental considerations. This approach could unveil how such information influences forecast accuracy.

Another potential avenue for future research is to consider a more detailed experimental setting to examine various measures of forecast accuracy, a subject that remains debated (Fildes et al., 2023). Given that different metrics may yield diverse results due to factors such as the size and granularity of historical data, experimental conditions (e.g., the number and demographics of subjects), and various industry domains (Davydenko & Fildes, 2016; Fildes & Goodwin, 2007), employing a complex experimental design with large sample sizes and different assigned roles or task decomposition (Lee & Siemsen, 2016) for subjects across various industry datasets could enhance the robustness and validity of our findings. Incorporating triangulation and benchmarking with multiple methods (e.g., AvgRelMAE and GMRAE) for measuring forecast accuracy would further this goal.

Furthermore, with the observed bias pertaining to service-level requirements, and the moderating effects of promotions, the critical question is how to address this matter in judgmental forecasting. Further research is needed to explore potential solutions, such as designing a platform or forecasting support system (FSS) that (1) visually informs demand forecasters about the utilization of their forecasts in supply chain decision-making, including demand planning and S&OP, and (2) ensures that demand forecasters are exposed to a personalized set of information that helps mitigate personal biases. Given the recent advancements in technology, the use of FSSs is growing, which introduces new managerial challenges such as adapting to a company's needs (Asimakopoulos & Dix, 2013), changing user expectations from the FSS, and managing increasing amounts of data and types of contextual information (Webby, O'Connor, & Edmundson, 2005). Although numerous studies have addressed the technological aspects and capabilities of FSSs, as well as their impact on statistical forecast accuracy (e.g., Fildes et al., 2006; Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Goodwin, Fildes, Lawrence, & Konstantinos, 2007), very few studies have investigated FSSs from an organizational perspective. This includes the interaction among FSS users and other stakeholders involved in forecasting, as well as strategies that improve the quality of judgmental adjustments. A recommended approach to enhance judgment utilization is to introduce restrictions (Goodwin, Fildes, Lawrence, & Stephens, 2011). A critical aspect of improving demand forecasting is to understand when manual adjustments are necessary, instead of relying solely on forecasts generated through statistical models or artificial intelligence. Given the pivotal role of loss functions and human interventions in demand planning, there may be a need to minimize or even eliminate the use of human judgment in demand forecasting by automating the incorporation of various events.

Data and code availability

The data and code necessary to reproduce the works presented in the paper is made available at https://doi.org/10.5281/zenodo.13942993. The numerical results presented in this paper were reproduced by the Editor-in-Chief on the 20th of October 2024.

Acknowledgment

The authors would like to thank Professors Enno Siemsen, Robert Fildes, Marcus O'Connor, and Michael Lawrence for their constructive feedback at various stages of this project. Special thanks to Dr. Meysam Arvan for his technical assistance throughout the project.

Appendix A

See Fig. A.1.

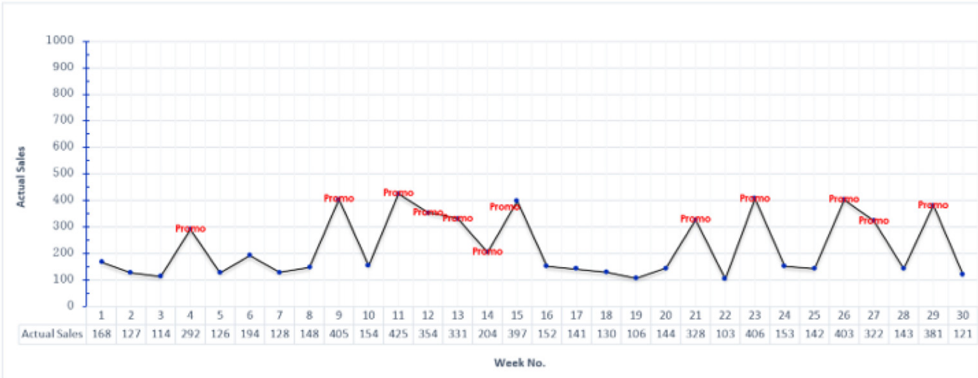
Appendix B. Performance of individual subjects across different treatments

Metrics: MAPE-P (mean absolute percentage deviation for promotional period), MAPE-NP (mean absolute percentage deviation for nonpromotional period), BIAS-P (forecast bias for promotional period), BIAS-NP (forecast bias for nonpromotional period).

Task Description

This section provides a description of the experiment. Please raise your hand at any point if you need further clarification.

Assume that you are a sales forecaster in a Fast Moving Consumer Goods (FMCG) company in Australia. Your task is to forecast demand of a popular product during the next four weeks. As a sales forecaster, your forecasts should only reflect the most likely value based on historical data and possible sales promotions. Your forecasts will then be forwarded to other departments (including demand planning and S&OP) where additional factors – such as internal/external constrains and requirements – will be taken into consideration to make related supply chain decisions. Figure below shows an example of what you will be seeing on your screen. The promotional periods are marked as "Promo". You will be notified whether the week you are forecasting for is a promotion or non-promotion week. You can assume that the impacts of promotions on the sales are independent from each other.



Available information:

- FMCG product.
- Long product shelf life (9 months).
- There is a **promotion** in week 31.

Your forecast:

Prior to starting the experiment, you will be asked to respond to a number of background questions. At the end of the experiment, you will be asked to state the factors that you considered when producing the forecasts.

Your performance will be assessed based on the average accuracy of the four forecasts you will make. The forecast accuracy is calculated using a mean absolute percentage error (MAPE) measure. All participants will receive an incentive payment of between $5 and $15 for participating in this experiment. The exact amount depends on your forecast accuracy; however, you will receive a minimum of $5 show-up fee.

Fig. A.1. An example of the task description presented to subjects.

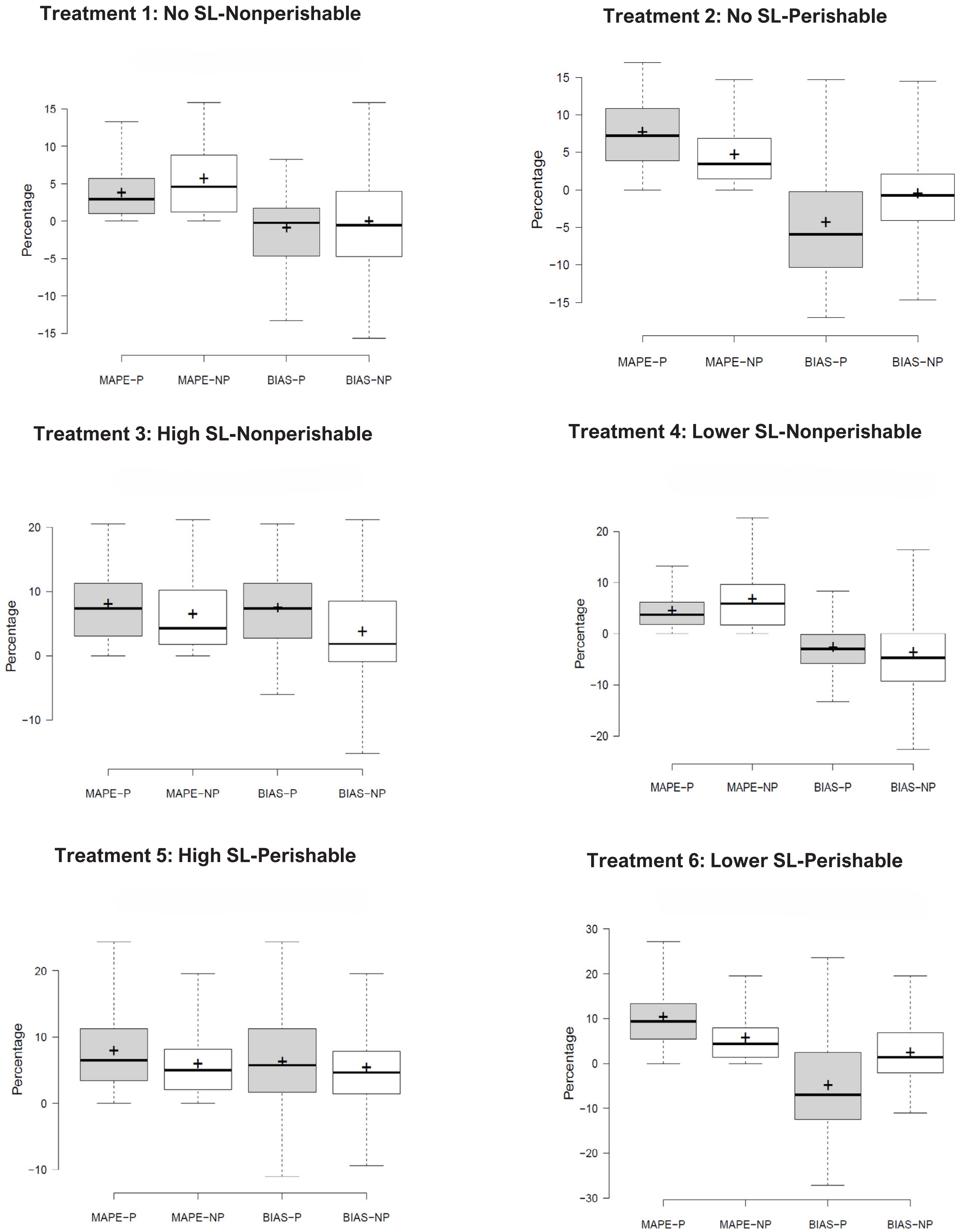


Fig. B.1. Comprehensive analysis of subject performance by treatments.

References

Abolghasemi, Mahdi, Hurley, Jason, Eshragh, Ali, & Fahimnia, Behnam (2020). Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. *International Journal of Production Economics*, [ISSN: 0925-5273] *230*, Article 107892. http://dx.doi.org/10.1016/j.ijpe.2020.107892, URL https://www.sciencedirect.com/science/article/pii/S0925527320302553.

Alvarado-Valencia, J. A., & Barrero, L. H. (2014). Reliance, trust and heuristics in judgmental forecasting. *Computers in Human Behavior*, *36*, 102–113.

Alvarado-Valencia, J., Barrero, L. H., Önkal, D., & Dennerlein, J. T. (2017). Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting. *International Journal of Forecasting*, *33*(1), 298–313.

Artto, K. A., & Pylkkänen, E. (1999). An effective procedure for the distribution of magazines. *International Transactions in Operational Research*, *6*(3), 289–310.

Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2018). Integrating human judgement into quantitative forecasting methods: A review. *Omega*, *86*, 237–252.

Asimakopoulos, S., & Dix, A. (2013). Forecasting support systems technologies-in-practice: A model of adoption and use for product forecasting. *International Journal of Forecasting*, *29*(2), 322–336.

Baecke, P., De Baets, S., & Vanderheyden, K. (2017). Investigating the added value of integrating human judgement into statistical demand forecasting systems. *International Journal of Production Economics*, *191*, 85–96.

Bendoly, E., Donohue, K., & Schultz, K. L. (2006). Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, *24*(6), 737–752.

Bolton, G. E., Ockenfels, A., & Thonemann, U. W. (2012). Managers and students as newsvendors. *Management Science*, *58*(12), 2225–2233.

Brau, Rebekah, Aloysius, John, & Siemsen, Enno (2023). Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics. *Journal of Operations Management*, *69*(6), 965–982. http://dx.doi.org/10.1002/joom.1257, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/joom.1257.

Broeke, M. Van. den., Baets, S. De., Vereecke, A., & Baecke, P. (2019). Judgmental forecast adjustments over different time horizons. *Omega*, *87*, 34–45.

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, *109*, 1997–2028.

Chen, F.-L., & Ou, T.-Y. (2009). Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store. *Expert Systems with Applications*, *36*(9), 7054–7063.

Craig, N., Dehoratius, N., & Raman, A. (2013). *The impact of supplier reliability on retailer demand*. Chicago: Harvard Business School.

Davydenko, A., & Fildes, R. (2016). Forecast error measures: Critical review and practical recommendations. In *Business forecasting: practical problems and solutions*. John Wiley & Sons Inc.

De Baets, Shari, & Harvey, Nigel (2018). Forecasting from time series subject to sporadic perturbations: Effectiveness of different types of forecasting support. *International Journal of Forecasting*, [ISSN: 0169-2070] *34*(2), 163–180. http://dx.doi.org/10.1016/j.ijforecast.2017.09.007, URL https://www.sciencedirect.com/science/article/pii/S0169207017301048.

De Baets, S., & Harvey, N. (2020a). Forecast adjustments over different horizons. *Decision Sciences*, *54*(4), 734–764.

De Baets, S., & Harvey, N. (2020b). Forecasting from time series subject to sporadic perturbations: Effectiveness of different types of forecasting support. *International Journal of Forecasting*, *34*(2), 163–180.

De Baets, S., & Harvey, N. (2020c). Judgment in predictive analytics. In M. Seifert (Ed.), *International series in operations research & management science*: *vol. 343*, (pp. 265–287). Springer.

De Baets, S., & Harvey, N. (2020d). Using judgment to select and adjust forecasts from statistical models. *European Journal of Operational Research*, *284*(1), 882–895.

Desanctis, Gerardine, & Jarvenpaa, Sirkka L. (1989). Graphical presentation of accounting data for financial forecasting: An experimental investigation. *Accounting, Organizations and Society*, [ISSN: 0361-3682] *14*(5), 509–525. http://dx.doi.org/10.1016/0361-3682(89)90015-9, URL https://www.sciencedirect.com/science/article/pii/0361368289900159.

Dickson, G. W., DeSanctis, G., & McBride, D. J. (1986). Understanding the effectiveness of computer graphics for decision support: A cumulative experimental approach. *Communications of the ACM*, *20*, 68–102.

Diermann, Christoph, & Huchzermeier, Arnd (2017). Case article–canyon bicycles: Judgmental demand forecasting in direct sales. *INFORMS Transactions on Education*, *17*(2), 58–62. http://dx.doi.org/10.1287/ited.2016.0165ca.

Dipboye, R. L., & Flanagan, M. F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? *American Psychologist*, *34*(2), 141–150.

DuHadway, S., & Dreyfus, D. (2017). A simulation for managing complexity in sales and operations planning decisions. *Decision Sciences Journal of Innovative Education*, *15*(4), 330–348.

ECR Australia (2010). Winning with promotions industry report.

Eiglsperger, J., Haselbeck, F., Stiele, V., Serrano, C. Guadarrama., Lim-Trinh, K., Menrad, K., et al. (2024). Forecasting seasonally fluctuating sales of perishable products in the horticultural industry. *Expert Systems with Applications*, *249*, Article 123438.

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133.

Eroglu, C., & Sanders, N. R. (2020). Effects of personality on the efficacy of judgmental adjustments of statistical forecasts. *Management Decision*, *60*(3), 589–605.

Fahimnia, B., Arvan, M., Tan, T., & Siemsen, E. (2023). A hidden anchor: The influence of service levels on demand forecasts. *Journal of Operations Management*, *69*(5), 856–871.

Fahimnia, Behnam, Pournader, Mehrdokht, Siemsen, Enno, Bendoly, Elliot, & Wang, Charles X. (2019). Behavioral operations and supply chain management-A review and literature mapping. *Decision Sciences*, *50*, 1127–1183, URL https://api.semanticscholar.org/CorpusID:159329484.

Fildes, R. (2006). The forecasting journals and their contribution to forecasting research: Citation analysis and expert opinion. *International Journal of Forecasting*, *22*, 415–432.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570–576.

Fildes, R., & Goodwin, P. (2021). Stability in the inefficient use of forecasting systems: A case study in a supply chain company. *International Journal of Forecasting*, *37*(4), 1031–1046.

Fildes, R., Goodwin, P., & De Baets, S. (2023). Forecast value added in demand planning. *SSRN Electronic Journal*.

Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, *42*(1), 351–361.

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., Goodwin, P., Önkal, D., & Lawrence, M. (2019). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, *35*(1), 144–156.

Fildes, R., Goodwin, P., Önkal, D., & Thomson, M. (2009). Effective use of forecasts in judgmental decisions: Principles and empirical evidence. *Decision Sciences*, *40*(3), 433–460.

Fildes, Robert, Ma, Shaohui, & Kolassa, Stephan (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, [ISSN: 0169-2070] *38*(4), 1283–1318, Special Issue: M5 competition. http://dx.doi.org/10.1016/j.ijforecast.2019.06.004 URL https://www.sciencedirect.com/science/article/pii/S016920701930192X.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288–299.

Franses, P. H. (2013). Improving judgmental adjustment of model-based forecasts. *Mathematics and Computers in Simulation*, *93*, 1–8.

Franses, P. H., & Legerstee, R. (2009). Properties of expert adjustments on model-based SKU-level forecasts. *International Journal of Forecasting*, *25*(1), 35–47.

Goodwin, Paul (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, [ISSN: 0169-2070] *16*(1), 85–99. http://dx.doi.org/10.1016/S0169-2070(99)00026-6, URL https://www.sciencedirect.com/science/article/pii/S0169207099000266.

Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, *12*(1), 37–53.

Goodwin, P., Fildes, R., Lawrence, M., & Konstantinos, N. (2007). The process of using a forecasting support system. *International Journal of Forecasting*, *23*(3), 391–404.

Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega*, *39*(3), 242–253.

Goodwin, P., & Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological Forecasting and Social Change*, *77*(3), 355–368.

Harvey, N. (2001). Improving judgment in forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting* (pp. 59–80). Springer Science+Business Media.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of presentation format on judgmental forecasting. *International Journal of Forecasting*, *12*(1), 119–137.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117–133.

Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment? *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(2), 589–607.

Heizer, J., Render, B., & Munson, C. (2017). *Operations management: sustainability and supply chain management* (12th ed.). (pp. 533–534). Pearson.

Hewage, H. C., Perera, H. N., & De Baets, S. (2022). Forecast adjustments during post-promotional periods. *European Journal of Operational Research*, *300*(1), 461–472.

Hofer, P., Eisl, C., & Mayr, A. (2015). Forecasting in Austrian companies: Do small and large Austrian companies differ in their forecasting processes? *International Journal of Operations and Production Management*, *9*(8), 40–62.

Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, *27*(2), 115–138.

Huber, Jakob, Gossmann, Alexander, & Stuckenschmidt, Heiner (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications*, *76*(1).

Ibrahim, Rouba, Kim, Song-Hee, & Tong, Jordan (2021). Eliciting human judgment for prediction algorithms. *Management Science*, *67*(4), 2314–2325. http://dx.doi.org/10.1287/mnsc.2020.3856.

Ivert, L. K., & Jonsson, P. (2010). The potential benefits of advanced planning and scheduling systems in sales and operations planning. *Industrial Management and Data Systems*, *110*, 659–681.

Kaipia, Riikka, Holmström, Jan, Småros, Johanna, & Rajala, Risto (2017). Information sharing for sales and operations planning: Contextualized solutions and mechanisms. *Journal of Operations Management*, [ISSN: 0272-6963] *52*, 15–29. http://dx.doi.org/10.1016/j.jom.2017.04.001, URL https://www.sciencedirect.com/science/article/pii/S0272696317300281.

Khosrowabadi, N., Hoberg, K., & Imdahl, C. (2022). Evaluating human behaviour in response to AI recommendations for judgemental forecasting. *European Journal of Operational Research*, *303*(1), 1151–1167.

Klassen, R. D., & Flores, B. E. (2001). Forecasting practices of Canadian firms: Survey results and comparisons. *International Journal of Production Economics*, *70*(2), 163–174.

Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, *57*(10), 1827–1843.

Langer, E. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311–328.

Lawrance, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493–518.

Lawrence, M. J. (1983). An exploration of some practical issues in the use of quantitative forecasting models. *Journal of Forecasting*, *2*(2), 169–179. http://dx.doi.org/10.1002/for.3980020207, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980020207.

Lawrence, Michael, O'Connor, Marcus, & Edmundson, Bob (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research*, [ISSN: 0377-2217] *122*(1), 151–160. http://dx.doi.org/10.1016/S0377-2217(99)00085-5, URL https://www.sciencedirect.com/science/article/pii/S0377221799000855.

Lee, C.-M. (2008). A Bayesian approach to determine the value of information in the newsboy problem. *International Journal of Production Economics*, *112*(1), 391–402.

Lee, Y. S., & Siemsen, E. (2016). Task decomposition and newsvendor decision making. *Management Science*, *63*(10), 3226–3245.

Lei, Dazhou, Hu, Hao, Geng, Dongyang, Zhang, Jianshen, Qi, Yongzhi, Liu, Sheng, et al. (2023). New product life cycle curve modeling and forecasting with product attributes and promotion: A Bayesian functional approach. *Production and Operations Management*, *32*(2), 655–673. http://dx.doi.org/10.1111/poms.13892, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/poms.13892.

Locke, E. A. (1986). Generalizing from laboratory to field settings: Research findings from industrial–organizational psychology, organizational behavior, and human resource management. *Industrial-Organizational Psychologist*, *33*(2), 57–64.

Lundqvist, J., De Fraiture, C., & Molden, D. (2008). *Saving water: from field to fork curbing losses and wastage in the food chain*. Stockholm: Stockholm International Water Institute.

Ma, Shaohui, & Fildes, Robert (2021). Retail sales forecasting with meta-learning. *European Journal of Operational Research*, [ISSN: 0377-2217] *288*(1), 111–128. http://dx.doi.org/10.1016/j.ejor.2020.05.038, URL https://www.sciencedirect.com/science/article/pii/S0377221720304847.

Ma, Shaohui, Fildes, Robert, & Huang, Tao (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*, [ISSN: 0377-2217] *249*(1), 245–257. http://dx.doi.org/10.1016/j.ejor.2015.08.029, URL https://www.sciencedirect.com/science/article/pii/S0377221715007845.

Mary E. Thomson, Karen B. Henriksen, & Macaulay, Alex (2004). The influence of the forecast horizon on judgemental probability forecasts of exchange rate movements. *The European Journal of Finance*, *10*(4), 290–307. http://dx.doi.org/10.1080/13518470110047620.

McCarthy, T. M., Davis, D. F., Golicic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, *25*(5), 303–324.

Mena, C., Adenso-Diaz, B., & Yurt, O. (2011). The causes of food waste in the supplier-retailer interface: Evidences from the UK and Spain. *Resources, Conservation and Recycling*, *55*(6), 648–658.

Morlock, H. (1967). The effect of outcome desirability on information required for decisions. *Behavioral Science*, *12*(4), 296–300.

Nakano, M. (2009). Collaborative forecasting and planning in supply chains, the impact on performance in japanese manufacturers. *International Journal of Physical Distribution and Logistical Management*, *39*, 84–105.

Nikolopoulos, K., Litsa, A., Petropoulos, F., Bougioukos, V., & Khammash, M. (2015). Relative performance of methods for forecasting special events. *Journal of Business Research*, *68*(12), 1785–1791.

Oliva, R., & Watson, N. (2011). Cross-functional alignment in supply chain planning: A case study of sales and operations planning. *Journal of Operations Management*, *29*(6), 434–448.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*, 390–409.

Perera, H. N., Hurley, J., Fahimnia, B., & Reisi, M. (2019). The human factor in supply chain forecasting: A systematic review. *European Journal of Operational Research*, *274*(2), 574–600.

Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behavior? *European Journal of Operational Research*, *249*(3), 842–852.

Petruzzi, N. C. (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research*, *47*(2), 183–194.

Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. H. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research*, *213*(2), 361–374.

Sanders, N. R., & Manrodt, K. B. (1994). Forecasting practices in US corporations: Survey results. *Interfaces*, *24*(2), 92–100.

Sanders, N. R., & Manrodt, K. B. (2003). Forecasting software in practice: Use, satisfaction and performance. *Interfaces*, *33*(5), 90–93.

Sanders, N. R., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, *5*, 39–52.

Sanders, N. R., & Ritzman, L. P. (2001a). Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting*. Norwell, MA: Kluwer Academic.

Sanders, N. R., & Ritzman, L. P. (2001b). Judgmental adjustment of statistical forecasts. In *Principles of forecasting* (pp. 405–416). Boston, MA: Springer.

Seifert, M., Siemsen, E., Hadida, A. L., & Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products. *Journal of Operations Management*, *36*, 33–45.

Šindelář, J. (2016). Investigation of factors influencing employee performance: A case of sales forecasting. *International Journal of Organizational Analysis*, *24*(2), 340–368.

Sodero, A. C. (2022). Good cause, not so good business? Sales and operations performance of cause-related marketing. *Journal of Business Logistics*, *43*, 566–592.

Srinivasan, S., Pauwels, K., Hanssens, D. M., & Dekimpe, M. G. (2004). Do promotions benefit manufacturers, retailers, or both? *Management Science*, *50*(5), 617–629.

Sroginis, A., Fildes, R., & Kourentzes, N. (2023). Use of contextual and model-based information in adjusting promotional forecasts. *European Journal of Operational Research*, *307*(1), 1177–1191.

Stewart, Thomas R. (2001). Improving reliability of judgmental forecasts. In J. Scott Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 81–106). Boston, MA: Springer US, ISBN: 978-0-306-47630-3, http://dx.doi.org/10.1007/978-0-306-47630-3_5.

Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: A 50-year review. *Journal of the Operational Research Society*, *60*(S1), S149–S160.

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, *26*(1), 134–143.

Thye, Shane R. (2014). Chapter 3 - Logical and philosophical foundations of experimental research in the social sciences. In Murray Webster, & Jane Sell (Eds.), *Laboratory experiments in the social sciences* (2nd ed.). (pp. 53–82). San Diego: Academic Press, ISBN: 978-0-12-404681-8, http://dx.doi.org/10.1016/B978-0-12-404681-8.00003-0, URL https://www.sciencedirect.com/science/article/pii/B9780124046818000030.

Tiwari, D., Patil, R., & Shah, J. (2011). Unreliable newsboy problem with a forecast update. *Operations Research Letters*, *39*(3), 278–282.

Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, *66*(2), 299–307.

Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, *29*(2), 234–243.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Van Donselaar, K. H., Peters, J., Jong, A. de., & Broekmeulen, R. A. C. M. (2016). Analysis and forecasting of demand during promotions for perishable items. *International Journal of Production Economics*, *172*, 65–75.

Webby, R., O'Connor, M., & Edmundson, B. (2005). Forecasting support systems for the incorporation of event information: An empirical investigation. *International Journal of Forecasting*, *21*(3), 411–423.

Webby, R., O'Connor, M., & LaWrence, M. (2001). Judgmental time-series forecasting using domain knowledge. In J. S. Armstrong (Ed.), *Principles of forecasting*. Norwell, MA: Kluwer Academic.