Int. J. Production Economics 278 (2024) 109442

Contents lists available at ScienceDirect

International Journal of Production Economics

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

Algorithm aversion during disruptions: The case of safety stock

Xun Wang [a,*], Vasco Sanchez Rodrigues [a], Emrah Demir [a], Joseph Sarkis [b]

[a] Cardiff Business School, Cardiff University, UK
[b] Foisei Business School, Worcester Polytechnic Institute, USA

A R T I C L E I N F O

Keywords:
Algorithm aversion
Inventory
Disruption
Simulation
Case study

A B S T R A C T

Algorithm aversion occurs when organizations or individuals reject optimal analytical decision support in favour of informal, subjective decisions. This phenomenon has been observed in many practical decision-making scenarios and is generally believed to negatively impact decision quality. However, its existence and effect in volatile supply chain environments has not been empirically tested in the literature. Safety stock buffering demand volatility is an important decision in supply chain management, making it an ideal lens to observe algorithm aversion. In this paper, we empirically investigate algorithm aversion behaviour in the context of safety stock settings. We collect data from a case retail company across a range of stockkeeping units (SKUs), encompassing both pre-disruption and post-disruption time stages with varying levels of volatility. We introduce a simulation model to determine whether algorithm aversion exists for safety stock decisions and to assess how algorithm adoption and adaptation affects performance. Our findings indicate that algorithm aversion occurs during supply chain disruptions, with algorithmic decisions significantly outperforming human judgment. Based on interview results and theories of information systems, we propose a theory to explain and generalize the above findings. This theory attributes algorithm aversion behaviour to reduced sense of fitness among algorithm users and lack of slack resources for both users and developers. It also offers insights into how the adoption and adaptation of algorithms influence decision performance during disruptive events.

1. Introduction

Algorithm aversion is a phenomenon where humans tend to trust their intuition more than analytical algorithms when making managerial decisions (Dietvorst et al., 2015). This aversion suggests that human decision-makers exhibit greater tolerance for errors made by human peers than for those made by computer codes and analytical models. Algorithm aversion has been observed in various contexts, including demand forecasting (Fildes et al., 2009), online sales recommendation (Yeomans et al., 2019), medical diagnosis and prescription (Longoni et al., 2019) and legal settings (Lowens, 2020). However, research on algorithm aversion behaviour in the context of operations management and supply chain management remains limited (Feng and Gao, 2020).

Supply chain resilience during disruptions has received considerable attention in recent years (Katsaliaki et al., 2022). By definition, supply chain disruptions refer to abrupt internal or external events that affect supply chain operations. Typical events triggering supply chain disruptions include natural disasters—such as earthquakes, volcanic eruptions, and pandemics—or human-made disruptions, such as economic crises, wars, and strikes (ibid.). A recent disaster illustrating the detrimental impact of supply chain disruptions is the COVID-19 pandemic. Disruptions like COVID-19 significantly increased supply chain risk. For instance, COVID-19 disruptions greatly increased the likelihood of supply chains experiencing the bullwhip effect and ripple effect (Scarpin et al., 2022). On the demand side, disruptions related to COVID-19 include increased demand, shortages, and disrupted demand patterns. These patterns have been some of the most significant supply chain challenges (Schleper et al., 2021; Seuring et al., 2022; Rinaldi et al., 2022). Supply chain risk during disruptions tends to follow a heteroskedastic pattern, meaning that the amplitude of the risk (e.g., forecast error) changes as the event unfolds over time (Browning et al., 2023).

High-quality decisions enable businesses to seize opportunities and address threats during disruptive events (Nikookar and Yanadori, 2022). Thus, given the increased frequency of disruptive events worldwide and the growing reliance on computer- and algorithm-based operational decision making, several critical questions emerge: Do decision-makers tend to avoid using algorithms during disruptive events? What are the

* Corresponding author.
E-mail address: wangx46@cardiff.ac.uk (X. Wang).

https://doi.org/10.1016/j.ijpe.2024.109442
Received 2 February 2024; Received in revised form 4 October 2024; Accepted 15 October 2024
Available online 18 October 2024

consequences of such behaviour? What factors drive this behaviour? Laboratory evidence suggests that humans exhibit algorithm aversion when the external environment is risky and volatile (Dietvorst and Bharti, 2020). However, this hypothesis has not been tested using actual supply chain scenarios and data, nor is there a comprehensive business or organizational theory to explain such behaviour. Understanding these questions is crucial because humans are the ultimate decision-makers in supply chains. This understanding can guide businesses in building robust data analytics capabilities and subtly influencing managerial decision-making processes. By doing so, businesses and supply chains can achieve better performance and enhance their resilience during disruptions.

Among the various types of supply chain decisions, the inventory replenishment decision—the determination of order quantity for individual SKUs (stockkeeping units)—is crucial in regulating physical and information flows. Decision support algorithms for calculating order quantity is widely adopted by manufacturing and retail businesses. Most of these inventory replenishment algorithms, such as the well-known order-up-to policy, are heuristic in nature (De Kok et al., 2018). In these algorithms, the order quantity for future periods is determined using inputs from various system functions, including demand and sales forecast, safety stock levels, and current stock levels (ibid.). The recommended order quantity is then calculated using these inputs.

Among these input variables, safety stock serves as a buffer to manage supply chain risks, particularly the risk of inaccurate demand forecasts (Krupp, 1997). The safety stock level is influenced by the decision-maker's prediction of future uncertainty (Gonçalves et al., 2020). Safety stock decisions become especially important during disruptive supply chain events, as uncertainty levels can fluctuate drastically in a short period (Christopher and Holweg, 2011). Thus, we choose this decision task to explore the phenomenon of algorithm aversion. Whether the decision-makers bypass algorithmic recommendations regarding safety stock levels reflects their trust in the algorithm and their assessment of risk.

We aim to address three key questions in this research study: (1) Is there algorithm aversion during supply chain disruptions? (2) Does algorithm aversion improve or worsen the performance of the algorithm? (3) Why do decision-makers exhibit algorithm aversion behaviour during disruptive events? To evaluate these questions, we utilise case study and simulation methodologies, along with a theoretical approach. The justification for these methodological approaches stems from limitations of previous studies. The case study method allows us to observe actual user behaviour recorded in the business database, ensures that the behaviours studied are realistic (Stuart et al., 2002). This method is particularly suitable for evaluating the first question. We considered a real-life case study where we collected operational data from a retail company for the year of 2020, during which many countries imposed full or partial lockdown measures in response to the COVID-19 pandemic. We also gathered data from a year before the disruption. This longitudinal case study approach, comparing pre- and post-disruption periods, provides insights into how algorithm aversion behaviour relates to changes in the disruptive environment.

Furthermore, simulation methodologies complement the case study approach in investigating the algorithm adoption and aversion behaviours. Under an experimental approach, the algorithmic performance is easy to observe or calculate, but the actual decision behaviour is unknown prior to the experiment (Castelo et al., 2019). Conversely, under a case study approach, reconstructing algorithmic recommendations and performance poses a significant methodological challenge. Additionally, the psychological constructs of *trust* and *tolerance*, which are key in behavioural research on algorithm aversion (Dietvorst et al., 2015), are difficult to retrospectively observe in real-world settings. Therefore, we need to shift our focus to adoption and adjustment behaviours—specifically, how much and how often actual decisions deviate from the algorithm's recommendations. Simulation is essential for evaluating *what-if* questions—for example, what happens if human decision makers strictly follow or alter the algorithmic decisions? This approach allows us to measure the performance of algorithmic recommendations under strict adherence, making it suitable for evaluating research question two.

This study contributes to the literature both normatively and empirically. First, we evaluate the existence of algorithm aversion during a real-life disruption, observed as intensified adjustments to algorithmic recommendations. These observations and evaluations complement the experimental evidence of algorithm aversion in the literature (Filiz et al., 2023). Second, using simulation, we show that algorithm aversion behaviour negatively impacts decision-making performance, building on findings from previous studies. Additionally, we demonstrate the superiority of alternative algorithms based on heteroskedasticity over the incumbent algorithm. This demonstration represents the first attempt to apply heteroskedastic models to inventory control and safety stock setting to counter supply chain disruptions. Lastly, we propose a theoretical framework to explain the existence and performance of algorithm aversion during disruptions, drawing from multiple theoretical perspectives. Based on this framework, we propose testable hypotheses for further validation, providing a foundation for future research to build upon.

This paper is organized as follows. In Section 2, we present the theoretical background of this research. Section 3 presents methodology including information on the case study and the data used for simulation. Section 4 gives the simulation results. Section 5 introduces a theoretical framework to explain the adoption and adaptation behaviours observed in this case study. Section 6 concludes the paper including study limitations and future research directions.

2. Literature review

The relevant literature of this research is reviewed in this section. This research contributes to the literature of algorithm aversion and inventory theory, in particular dynamic safety stock methods.

2.1. Algorithm aversion

Dietvorst et al. (2015) first introduced this term "algorithm aversion". The authors found that decision-makers exhibit lower tolerance towards algorithmic errors compared to human errors. This finding was supported by additional studies (e.g. Prahl and Van Swol, 2017). A comprehensive review of algorithm aversion (Mahmud et al., 2022) revealed that most existing studies on this phenomenon rely on experimentation. Algorithm aversion may also manifest as more frequent adjustments to algorithm recommendations. For instance, Fildes and Goodwin (2021) observed that managers often make frequent adjustments to algorithmic recommendations when making forecasts, resulting in reduced forecast accuracy and increased management burden.

Several proposed explanations exist for algorithmic aversion behaviour. Dietvorst and Bharti (2020) attributed this phenomenon to the diminishing sensitivity of decision-makers to forecast errors. Task mismatch has also been used as an antecedent factor for algorithm aversion (Lowens, 2020). Task mismatch occurs when decision-makers perceive the task, such as subjective evaluations, to be beyond the capabilities of the algorithm, leading them to be more averse to its use. Additionally, there are propositions that algorithm aversion is influenced by bias in evaluating others—both users and their peers—rather than being solely due to the algorithm (Morewedge, 2022). These contextual factors suggest that algorithm aversion is more likely to occur if the task is central to the user's identity or if performance measures are more ambiguous and subjective.

Studies have also explored methods to mitigate managerial algorithm aversion, under the assumption that aversive behaviour reduces decision quality. One approach involves providing social proof of algorithm superiority, which has been shown to effectively persuade people to adopt algorithms (Alexander et al., 2018). Alternatively, implementing policies that limit decision maker adjustments can increase the propensity to use algorithms without compromising performance (Dietvorst et al., 2018).

The primary approach in algorithm aversion studies has predominantly been experimental. One of the few exceptions focused on forecast adjustments (Fildes and Goodwin, 2021). In this study, algorithmic forecasts were generated by inputting actual data into the forecasting system. And their accuracy was compared with judgmental or judgmentally adjusted forecasts. We have adopted a similar simulation-based approach in our study, but with some modifications to enhance analysis. Firstly, our simulation model is customized for each individual SKU, considering their unique operational characteristics such as shipping lead-time and shelf-life. In contrast, forecasting algorithms are typically assumed to be applicable to all SKUs, without accounting for their specific characteristics. Secondly, our study diverges from the main objective of forecasting research, which typically focuses on accuracy. In inventory control, multiple objectives of similar or equal importance, such as availability, stock level, and waste, must be considered (see Section 3.4).

Several gaps exist in the current literature on the topic of algorithm aversion. First, existing studies on algorithm aversion in volatile environments predominantly focus on psychological dimensions of decision-makers, neglecting important managerial and organizational factors (Dietvorst and Bharti, 2020). Second, the algorithm aversion literature lacks investigation into an important operations and supply chain management decision—the inventory replenishment decision. Third, most research on algorithm aversion is conducted in laboratory settings, where subjects make decisions under controlled conditions. Actual empirical evaluation is missing, which is important to capture the complexities and imperfections of practical settings.

2.2. Risk, heteroskedasticity, and safety stock

In supply chain disruptive events, the level of risk can fluctuate over time (Christopher and Holweg, 2011). Heteroskedasticity, which involves time series with updating conditional variance, is a model that accounts for the time-varying volatility. During supply chain disruptions, time series of forecast errors can exhibit heteroskedasticity. The GARCH model (Generalized Auto-Regressive Conditional Heteroskedasticity, Bollerslev, 1986) is commonly used for modelling and predicting heteroskedastic time series. While the GARCH model has been applied in demand forecasting, it typically models the demand series itself, rather than the forecast error, to be heteroskedastic (e.g. see Zhang, 2007).

When the forecast error exhibits heteroskedasticity, dynamic safety stock algorithms tend to outperform constant or static algorithms, given that safety stock serves as a buffer to hedge against risk (Kanet et al., 2010; Stößlein et al., 2014). As the level of risk fluctuates over time, safety stock buffer should change accordingly. However, previously developed models aimed at optimizing safety stock in such environments often rely on simplifying assumptions and overlook inventory update processes. Some studies have addressed this issue (see e.g., Trapero et al., 2019) by employing GARCH for setting safety stock, similar to our research design. Although the GARCH model has demonstrated strong performance, these studies typically focus on general demand time series without explicitly considering disruptive events. For a comprehensive review of safety stock setting methods and techniques, including their advantages and limitations, see Gonçalves et al. (2020) and Barros et al. (2021).

Decision-maker attitude towards risk significantly influences safety stock setting given that safety stock serves to hedge against demand and supply risks. Individuals have been found to typically be risk averse, preferring small but certain payoffs over large but uncertain ones (Kanheman and Tversky, 1979). Although some differences may exist. For example, MacCrimmon and Wehrung (1990) surveyed over 500 top-level business executives and discovered that successful executives, in terms of wealth, position, and income, tend to be more risk-seeking, whereas mature executives, in terms of age, seniority, and the number of dependents, tend to be more risk-averse. Also, research suggests that female decision-makers exhibit greater risk aversion compared to their male counterparts (Eckel and Grossman, 2008).

Contradictory evidence exists regarding how risk propensity influences ordering behaviour and safety stock levels. In the context of newsvendor model, Eeckhoudt et al. (1995) analytically demonstrated that risk-averse newsvendors should decrease their order quantity, as larger orders lead to greater variability in payoff, regardless of the product's profit margin. This finding was experimentally confirmed by Becker-Peth et al. (2018). However, alternative perspectives suggest the opposite effect, indicating that risk aversion is associated with larger order sizes and higher inventory levels. Corbett and Fransoo (2007) conducted a survey among small business owners and entrepreneurs, revealing that subjects exhibit risk aversion toward gains but are risk-seeking toward losses, consistent with prospect theory. Moreover, they found that risk aversion is associated with higher inventory levels, particularly for high-margin products. Similarly, Cannella et al. (2019) observed a positive association between risk aversion and increased inventory level in a Beer Game experiment. These conflicting findings may stem from variations in the measurement of risk attitude. For instance, Becker-Peth et al. (2018) employed the incentivized Holt-Laury lottery, while Cannella et al. (2019) utilized a non-incentivized psychological inventory.

The research gap in the safety stock literature can be summarized as follows. While several dynamic safety stock algorithms based on techniques such as exponential smoothing or GARCH have been developed, their performance during supply chain disruptions have not been empirically evaluated. Furthermore, there has been limited investigation into user decision behaviour in the presence of safety stock algorithms. Existing research on the relationship between risk attitudes and order behaviour predominantly relies on experimental and interview approaches, lacking empirical support from actual operational data.

3. The case study and methodology

In this section, we present an overview of the case data and sources. The safety stock methods, the inventory simulation model, and the performance metrics adopted in the simulation are also defined.

3.1. The case and data overview

The case study conducted in this research focuses on a distribution centre (DC) operated by company X, a prominent online grocery retailer located in Europe. Company X offers a wide range of products, including perishable and non-perishable items such as food, beverages, personal care, and home care products. Each DC in company X's distribution network serves a specific geographical zone and is responsible for making replenishment and inventory management decisions independently. While some lateral transhipment occurs between DCs, it is negligible and can be disregarded for analysis purposes. Suppliers deliver products directly to the DCs, with most suppliers located within regional distances, resulting in short replenishment lead-times. However, some suppliers from neighbouring countries may have longer lead-times. Additionally, certain suppliers require orders to be placed in batch sizes, especially for non-perishable products, meaning that the order quantity must be a multiple of the specified batch size.

The company has implemented a demand management system comprising forecast models, an order assistance system, and order quantity recommendations. Daily, the forecasting system generates demand forecasts for each SKU for the next 30 days. The system utilizes both time-series and regressive forecasting methods. However, optimizing forecasting methods is not within the scope of this paper. The order recommendation system employs an *order-up-to* algorithm, which determines the recommended order quantity as the difference between the order-up-to level and the current inventory position. The order-up-to level is calculated as the sum of the lead-time demand forecast and the safety stock. The lead-time demand forecast is generated by the forecasting system, while the safety stock is determined as a multiple of the standard deviation of historical forecast errors. Further details on this algorithm are provided in Section 3.3.

Purchasing managers can manually adjust input variables to modify the order quantity based on their judgment. However, we have observed variations in the difficulty and complexity of adjusting these variables. The demand forecasts, generated by an independent forecasting system employing multiple forecasting algorithms, are relatively rigid and challenging for users to manually modify. Moreover, adjustments to demand forecasts primarily focus on enhancing forecast accuracy rather than optimizing inventory control. Stock level information, derived from inbound and outbound quantities or recorded directly from periodical inventory reviews, presents additional challenges for adjustments due to its utilization in auditing and financial reporting processes. In contrast, while an algorithm is employed to determine safety stock, purchasing managers typically possess greater flexibility in adjusting safety stock levels. By increasing or decreasing the safety stock, they can modify the recommended order quantity generated by the algorithm, allowing for deviations from algorithmic recommendations in the final purchasing decision.

The data utilized in this study comprises two datasets, each consisting of daily entries. The first dataset span from October 1, 2018 to September 30, 2019, while the second dataset covers the entire year of 2020, from 1 January to 31 December. Both datasets encompass one year of operational data, ensuring comparability in size. The dataset for 2020, reflecting the effects of the COVID lockdown disruption, is referred to as the *post-disruption* period (POST), while the dataset for the *pre-disruption* period is labelled as PRE.

The PRE dataset comprises 17 SKUs of fresh vegetables and fruits, while the POST dataset comprises 28 SKUs encompassing various food and grocery products such fruits, vegetables, raw meat, beverages, and personal hygiene items. These datasets were compiled from multiple attempts to extract data from the company's information system, ensuring comparability in terms of product category and characteristics. Recorded daily, the data fields include sales volume, order quantity, inventory level, received quantity and demand forecast. To mitigate scaling issues and preserve confidentiality, all data fields are normalized by dividing the annual average sales volume of each SKU over the year.

Fixed operational data for each SKU, including transportation lead-time and shelf life, is also collected. This data is relatively stable and with few, if any, changes. Hence, it is assumed that both the lead-time and shelf life remain constant. The transportation lead-time (referred to as lead-time) is defined as the time difference between when an order is placed to a supplier and when the products are received, measured in days. SKU lead-times range from one day to 14 days. Shelf life refers to the number of days an item can remain in the DC since arrival and ranges from two days to infinity, assumed for non-perishable products.

The dataset cannot be made publicly available due to confidentiality considerations.

3.2. Estimating heteroskedasticity

We commence with a descriptive analysis of the demand risk, specifically concentrating on evaluating variational shifts. Demand risk is quantified by the forecast error of sales during the lead-time period. Fig. 1 illustrates the daily fluctuations in demand forecast error for two example SKUs during the PRE and POST periods. Clearly, heteroskedasticity in forecast error becomes more pronounced following the disruption, with volatility seemingly aligning with the spread of the pandemic.

Given these characteristics, we apply Engle's ARCH test to the lead-time demand forecast error of all SKUs in both the PRE and POST datasets. Higher Engle's ARCH test statistic values indicate greater heteroskedasticity. As shown in Fig. 2, the POST dataset shows significant heteroskedasticity.

These initial observations inspired us to simulate algorithms for estimating the future variance of the forecast error, which is used for dynamically setting the safety stock. Three algorithms are simulated, including HIST, SES and GARCH. HIST uses historically observed forecast error variance as the estimation; SES estimates the variance with an exponential smoothing approach between the historical estimate and the newly observed forecast error; under GARCH, the forecast error series is modelled as a GARCH process, and the variance estimate is generated accordingly. The equations for these algorithms are provided

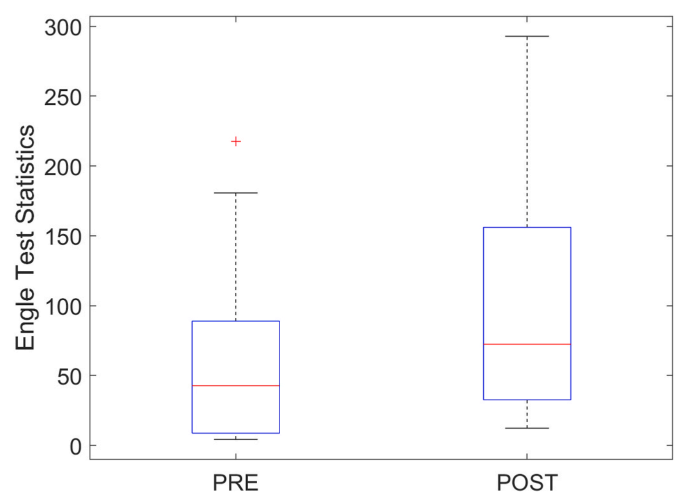

Fig. 2. Engle test statistics of forecast error in PRE and POST datasets.

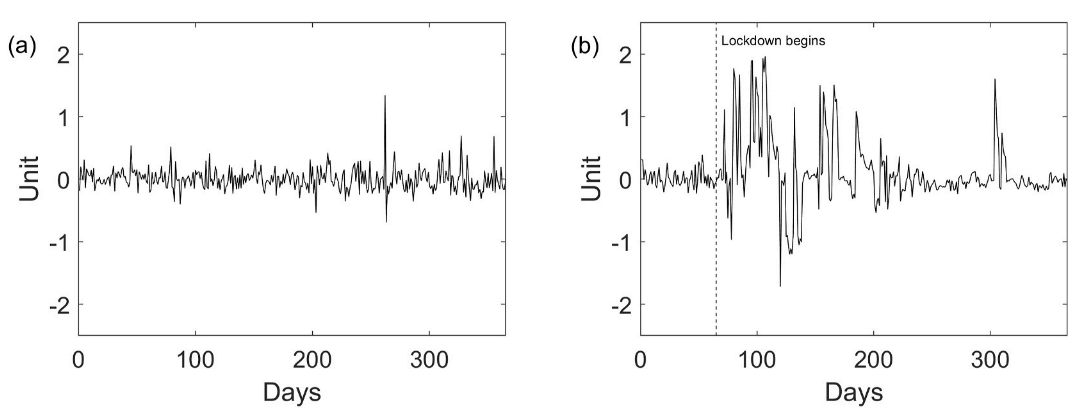

Fig. 1. The standardized forecast error of sample SKUs in PRE and POST. (a) SKU #6 in PRE; (b) SKU #26 in POST.

later in this section. The results of these algorithms are compared to the original order quantities (ORIG)—which are the actual orders made to the suppliers, as provided in the corporate datasets. It is important to note that although the SES and GARCH algorithms can be applied to predict the demand series, they are exclusively used to estimate the forecast error variance in this research. The demand forecasts are included in the datasets, and the discussion of demand forecasting methods is beyond the scope of this research.

The algorithms differ in estimating the future standard deviation of the lead-time demand forecast error, $\widehat{\sigma}_{t+1}$, which is used to set the safety stock. The time arrow assumption is not violated in the simulation—that is, all the information used in the forecasts and decisions are available prior to the time of evaluation.

The first algorithm is based on the historical forecast error (HIST), is to set the safety stock based on the historical forecast error, e.g., the past 30 days. It can be represented as (1):

$$\widehat{\sigma}^2_{t+1} = \frac{1}{30}\sum_{k=t-29}^{t}(\varepsilon_k - \overline{\varepsilon}_k)^2 \tag{1}$$

Let d_t be the sales of day t, and $\widehat{d}_{t,k}$ the sales forecast made at period k for period t. Therefore, $\sum_{k=1}^{L} d_{t-k}$ is the total sales over lead-time L and $\sum_{k=1}^{L} \widehat{d}_{t-k,t-L-1}$ is the sum of forecast made on period $t-L-1$ for the same span. $\varepsilon_t = \sum_{k=1}^{L} d_{t-k} - \sum_{k=1}^{L} \widehat{d}_{t-k,t-L-1}$ is the lead-time forecast error calculated at period t, and $\overline{\varepsilon}_t = \left(\sum_{k=t-29}^{t}\varepsilon_k\right)/30$ is the average forecast error over the last 30 periods.

Although the HIST method can detect long-term increasing and decreasing volatility trends, there is a lag between the actual and estimated change in volatility. We conjecture this algorithm is the incumbent safety stock policy currently adopted in the system. The evidence comes from two sources. First, the description of the interviewee of the incumbent policy matches a HIST policy. Second, we plot the order quantities recommended by HIST against the original quantities (see Fig. 3). HIST aligns with the original quantities better in PRE than in POST, suggesting that less adjustment is made during the PRE period. Meanwhile in POST, the diagonal fitting line shows an obvious upward bias, so much so that R^2 becomes negative. This can be attributed to the positive adjustment to the recommended order quantity.

The second algorithm (SES) is simple exponential smoothing which is used to predict the variance of forecast error. In other words, the variance of the forecast error in the next period equals a weighted average of the estimated forecast error variance in the current period and the square of current forecast error as in expression (2):

$$\widehat{\sigma}^2_{t+1} = (1-\beta)\widehat{\sigma}^2_t + \beta\varepsilon^2_t \tag{2}$$

β is the smoothing parameter defined by the user. This method has been proposed by Bretschneider (1986) due to its simplicity. Boudoukh et al. (1997) discussed the use of SES and GARCH to forecast volatilities in financial data. They found that SES outperforms GARCH. In this research, we automatically update the β value in each period such that it is optimal to minimize the forecast error up to the current period under the SES method—equation set (3). Let β_T be the β value to be used in period T, then:

$$\begin{aligned} &\beta_T = \operatorname{argmin} \sum_{t=1}^{T}\varepsilon_t^2 \\ &\text{s.t.}\quad \varepsilon_t = d_t - \widehat{d}_{t-L,L} \\ &\qquad \widehat{d}_{t,L} = \beta_T d_{t-1} + (1-\beta_T)\widehat{d}_{t-1,t} \end{aligned} \tag{3}$$

$\widehat{d}_{t,k}$ is the demand forecast made at t for k periods later. The updating mechanism ensures that the β value for estimating the variance in every period is the retrospect optimal value for the L-step-ahead forecast in the past T days. However, we have found that using an arbitrarily constant smoothing value β, for instance, $\beta = 0.5$, generates comparable performance.

Lastly, we adopt the GARCH model to predict the future changes in variability. The GARCH model is suitable for the forecasting of heteroskedastic time-series. The general GARCH (p, q) model is given as follows. Let the forecast error be $\{\varepsilon_t\}$, with $\varepsilon_t|\vartheta_{t-1}$ following a normal distribution with changing variance, $N(0,\sigma_t^2)$, where ϑ_{t-1} is the set of all available information up until $t-1$. The variance σ_t^2 follows the autoregressive process (4):

$$\sigma_t^2 = \omega + \varphi_1\varepsilon_{t-1} + \cdots + \varphi_q\varepsilon_{t-q} + \theta_1\sigma^2_{t-1} + \cdots + \theta_p\sigma^2_{t-p} \tag{4}$$

In this study, we adopt the GARCH(1,1) model and automatically update the φ_1 and θ_1 values using the estimate function in MATLAB R2019 (MATLAB, 2018).

As an example, Fig. 4 shows the safety stock generated by the four estimation methods, along with the demand series for SKU #26 post-COVID. It can be seen how the ORIG estimation method exaggerates the variance of forecast error, hence the safety stock.

3.3. The inventory model

We adopt the order-up-to policy as the ordering policy in the system model. This policy is widely adopted in both academia and industry due to its simplicity. This choice is also consistent with the case company. We will show here that under this policy, the safety stock should depend on the standard deviation of the forecast error. The linear order-up-to policy can be expressed as follows (5):

$$o_t = \sum_{k=1}^{L}\widehat{d}_{t,k} + ss_t - ip_t \tag{5}$$

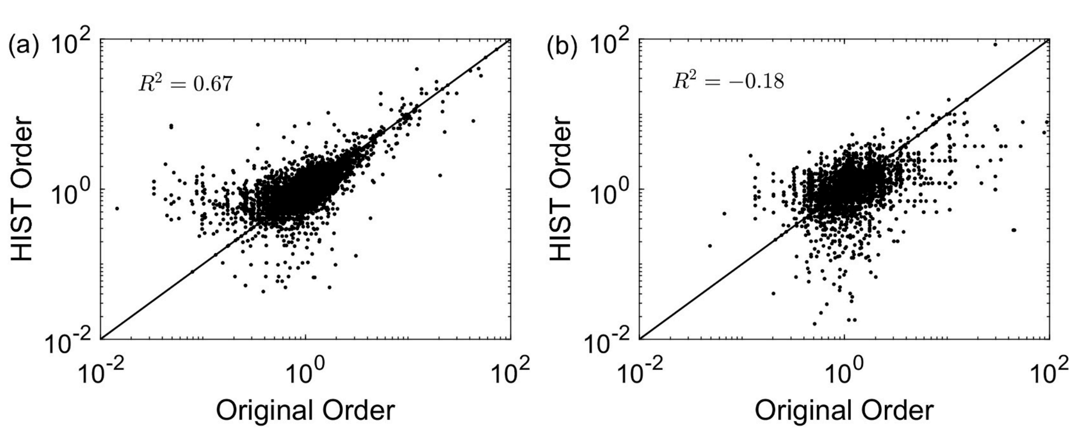

Fig. 3. Goodness-of-fit of HIST algorithm using logarithmic scales. (a) PRE; (b) POST.

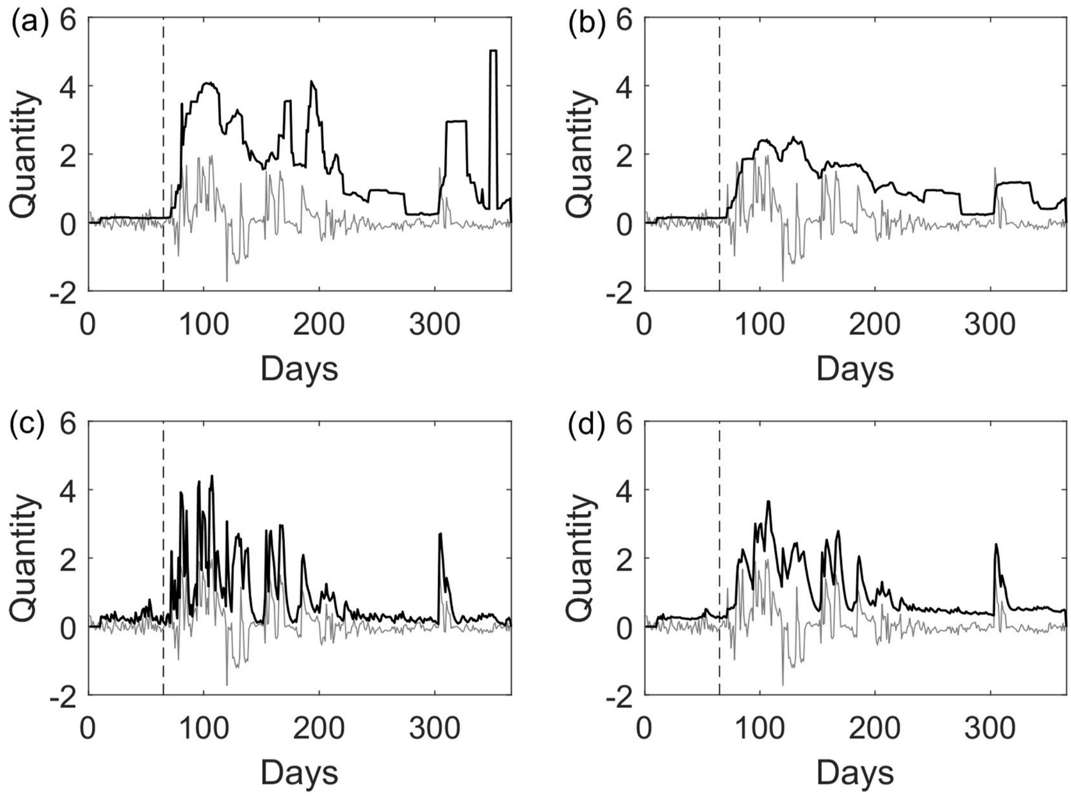

Fig. 4. Safety stock levels for SKU #26. (a) ORIG; (b) HIST; (c) SES; (d) GARCH.

$\sum_{k=1}^{L} \widehat{d}_{t,k}$ is the lead-time demand forecast. ss_t is the safety stock. Due to the heteroskedastic nature of demand and supply risk, the safety stock is time varying. ip_t is the inventory position, which equals the sum of inventory level, i_t, and work-in-process level, w_t. They satisfy the balance equation (6)(7):

$$i_t = i_{t-1} + c_t - d_t \tag{6}$$

and

$$w_t = w_{t-1} + o_{t-1} - c_t, \tag{7}$$

where L is the lead-time (the time difference between placing and receiving an order) which is assumed to be constant in our study. This is a valid assumption based on our observations of the case company, as the lead-time did not change during the disruptive period. c_t is the completed (arrived) quantity. If the supply-side uncertainty risk is absent, then $c_t = o_{t-L}$. The work-in-process can also be expressed as the sum of incoming orders, $w_t = \sum_{k=1}^{L-1} o_{t-k}$. The inventory position thus satisfies the balancing relationship (8):

$$ip_t = ip_{t-1} + o_{t-1} - d_t. \tag{8}$$

Equation (5) can be expressed verbally as "the order quantity equals the forecast of lead-time demand plus the safety stock minus the inventory position". It can be further rewritten as $o_t = \widehat{d}_{t,L} + ss_t - \left(i_t + \sum_{k=1}^{L-1} o_{t-k} - \sum_{k=1}^{L-1} \widehat{d}_{t,k}\right)$, where $\widehat{i}_{t,L} = i_t + \sum_{k=1}^{L-1} o_{t-k} - \sum_{k=1}^{L-1} \widehat{d}_{t,k}$ is the forecasted inventory level at $t + L$ based on the current inventory level at t, the incoming orders and the forecasted demand. The order-up-to policy can then be expressed as "the order quantity equals to the forecast demand minus the forecasted inventory, plus the safety stock".

On the other hand, from Equation (5), we have $i_t = \sum_{k=1}^{L} \widehat{d}_{t,k} - \sum_{k=1}^{L} o_{t-L+k} + ss_t$. From (5), we also have $i_{t+L} = i_t + \sum_{k=1}^{L} o_{t-L+k} - \sum_{k=1}^{L} d_{t+k}$. Combining these two equations, we have

$$i_{t+L} = ss_t + \sum_{k=1}^{L} \widehat{d}_{t,k} - \sum_{k=1}^{L} d_{t+k} \tag{9}$$

Equation (9) means that the inventory level equals to the safety stock minus the forecast error of the lead-time demand. Therefore, the optimal safety stock should be set in a newsvendor fashion based on the estimate of the standard deviation of lead-time demand forecast error $\widehat{\sigma}_{t+L} = \text{std}\left(\sum_{k=1}^{L} \widehat{d}_{t,k} - \sum_{k=1}^{L} d_{t+k}\right)$. This is the L-step-ahead estimate since only the forecast error up until the current period, $\sum_{k=1}^{L} \widehat{d}_{t-L,k} - \sum_{k=1}^{L} d_{t-L+k}$, can be observed. This estimate can be generated with the HIST, SES and GARCH algorithms given in Section 3.2. It should be noted that under HIST and SES methods, the estimate of $\widehat{\sigma}_{t+L}$ does not change with L. The safety stock is then set as $ss_t = q_\alpha \widehat{\sigma}_{t+L}$, where q_α is the α-quantile of standard normal distribution, and α is the target availability (Beutel and Minner, 2012; also see Section 3.4).

There are several complicating factors in real operations that may affect the validity of the above result. First, the order quantity o_t and the inventory level i_t cannot take negative values in practice because the retail business cannot return excessive inventory to the supplier, nor can they keep the undelivered customer orders until the product becomes available. These situations are referred to as "forbidden returns" and "lost sales" in the inventory management literature. To represent these situations mathematically, the right-hand side of (5) and (6) should be wrapped with the nonnegative constraint max $\{\bullet, 0\}$.

Second, the inventory balance equation (8) can only be applied to non-perishable products or those with a sufficiently long life. For perishable products with a short life, we use Nahmias (1982)'s perishable inventory model to simulate the inventory update process. This model uses the inventory compartmentalization technique, where inventory is divided into parts (compartments) according to the remaining life. Each compartment updates via a unique balance equation. Let $\widehat{m}$ be the largest remaining life, and $\check{m}$ as the smallest remaining life. Denote i_t^m as the inventory level with m days remaining life at the end of t, then it follows the following balance equations:

$$\begin{aligned} i_t^m &= \max\left[i_{t-1}^{m+1} - \max\left(d_t - \sum_{k=1}^{m} i_{t-1}^{k}, 0\right), 0\right], \text{for } m < \widehat{m} \\ i_t^m &= \max\left[o_{t-L} - \max\left(d_t - \sum_{k=1}^{m} i_{t-1}^{k}, 0\right), 0\right], \text{for } m = \widehat{m} \end{aligned} \tag{10}$$

We offer an intuitive explanation of the above equation (10) as follows: the demand (d_t) will be satisfied by the inventory older than $m+1$ first. The unsatisfied demand will be then satisfied by the inventory with the remaining life of $m+1$. Any leftovers will become inventory with remaining life of m in the next period. The amount of waste due to expiration is thus (11):

$$v_t = \max\left(i_{t-1}^{\bar{m}-d_t,0}\right), \tag{11}$$

which means that if the shortest life inventory is not sold, it expires.

Equation (10) assumes a first-in-first-out (FIFO) assumption, where the retailer will distribute the oldest items to consumers before they start to distribute newer ones. The validity of this assumption is greater for online retailers, as it is easier to achieve FIFO when the picking and distribution is managed centrally ((Nahmias, 1982); Barto et al., 2024). For brick-and-mortar stores, various complicating factors exist, such as filling the shelf from the DC, and consumers' picking behaviour at the shelf.

When these complicating factors exist, the optimal safety stock is no longer a multiple of the standard deviation of the forecast error. Nonetheless, the relationship still holds that the larger the variance of the forecast error, the higher the safety stock should be. This assumption is sufficient for our analysis, which focuses on decision maker perception and prediction about the demand uncertainty (variance of forecast error) and their response.

Finally, we make a note on supply uncertainty—defined as the phenomenon that the received quantity at the retailer is different (usually smaller) than the ordered quantity and the difference is uncertain. Supply uncertainty is an important component of supply chain risks, and we have observed the increase of supply uncertainty during the pandemic. Although the supply uncertainty is not the focus of this research, we nonetheless include it in our simulation model. Specifically, we assume that the received quantity of a particular order is proportional to the order quantity, and the proportion is determined in the original dataset:

$$\frac{c_{t+L}^{\text{alg}}}{o_t^{\text{alg}}} = \frac{c_{t+L}^{\text{orig}}}{o_t^{\text{orig}}} \tag{12}$$

The superscript alg represents the respective quantities based on the alternative algorithms, while orig represents the original quantity included in the dataset. This assumption holds if the supplier adopts the *linear allocation policy* to allocate the limited supply across retailers (Cachon and Lariviere, 1999). This assumption ensures that all candidate strategies suffer from the same level of supply uncertainty.

3.4. Performance metrics

We adopt three performance measures which are all commonly used in the inventory control practices. The first measure (and perhaps the most important one in retail) is product availability (α_c) defined as the proportion of periods in which no shortage occurs. In our simulation, it is calculated as the proportion of days when the inventory level after demand fulfilment stays above zero, since a positive inventory level indicates no shortage.

$$\alpha_c = \frac{\sum_{t=1}^{T} \mathbf{1}\{i_t > 0\}}{T} \tag{13}$$

In equation (13), $\mathbf{1}\{\bullet\}$ is an indicator function which equals to one when the argument is positive, and zero otherwise. Availability affects retailer profit and goodwill, hence an important metric for inventory management. The second measure is on-hand inventory level after demand fulfilment ($\bar{i}$) averaged over days. It increases with availability since a high inventory level reduces the likelihood of shortage (14).

$$\bar{i} = \frac{\sum_{t=1}^{T} i_t}{T} \tag{14}$$

The inventory level metric is important due to holding cost. In practice, high DC utilization forces the retailer to exploit other solutions, such as external storage, resulting in additional holding costs. As the actual inventory data is normalized by the average annual sales, the measure obtained from (14) should be interpreted as the average inventory coverage, i.e., the number of days that the inventory can cover the demand.

The final performance measure is average daily waste generated—defined as unsold inventory with a shelf life of zero (15).

$$\bar{v} = \frac{\sum_{t=1}^{T} v_t}{T} \tag{15}$$

This metric directly relates to food waste—given that most perishable products are within the food category. This metric has economic, social, and environmental consequences. Food waste reduces the profit of the retailer, general social food availability, and unnecessarily increases natural resources use (material and energy) in production and distribution.

When calculating these metrics, we also note that for both PRE and POST datasets, we exclude the initial 30 days from the reporting to reduce initialization effects.

4. Simulated inventory performance

In this section we present the simulation results, summarized in Fig. 5 and Table 1. The result from the PRE dataset is colour coded as grey and the POST dataset as black. The subfigures 5(a), 5(b) and 5(c) show achieved availability, inventory level, and waste, respectively. The bars represent the 95% confidence interval across all SKUs in the respective datasets. We compare the performance of the actual order quantity decisions with judgmental adjustments (ORIG), and the three algorithms introduced in Section 3.2 (HIST, SES and GARCH).

In looking at Fig. 5(a) we see that before disruption the benefit of algorithms to availability is not significant. The one-way ANOVA analysis across the four solutions shows no significant difference between any pair, with $F(3,64) = 0.16, p = 0.92$. This result is an indication of homoskedasticity in the pre-disruption era. Hence, the benefit of algorithms is not obvious.

After disruption takes place, each algorithm led to significant availability performance improvement when compared to the baseline (original) solution, $F(3, 108) = 10.42$, $p < 0.001$. Paired t-tests for comparisons between the policies appears in Table 1. We observe that the availability increases by 10.6% if the safety stock settings follow the current algorithm (HIST), and 16.2% if the SES and GARCH methods are adopted. Moreover, the SES and GARCH show superiority to HIST, with availability increases 5.1%. There is no significant difference between SES and GARCH.

The ANOVA analysis shows that the algorithm adoption and adaptation have a significant impact on the average inventory level performance in the PRE period ($F = 9.91; p < 0.001$), but not the POST period ($F = 0.52; p = 0.67$). Pairwise, in the PRE period, algorithm adaptation (SES vs. HIST) can reduce the inventory level, to an extent of 39.2%, compared with the original judgment. Likewise, GARCH can reduce the inventory level to a lesser extent (27%). In the POST period, the algo-

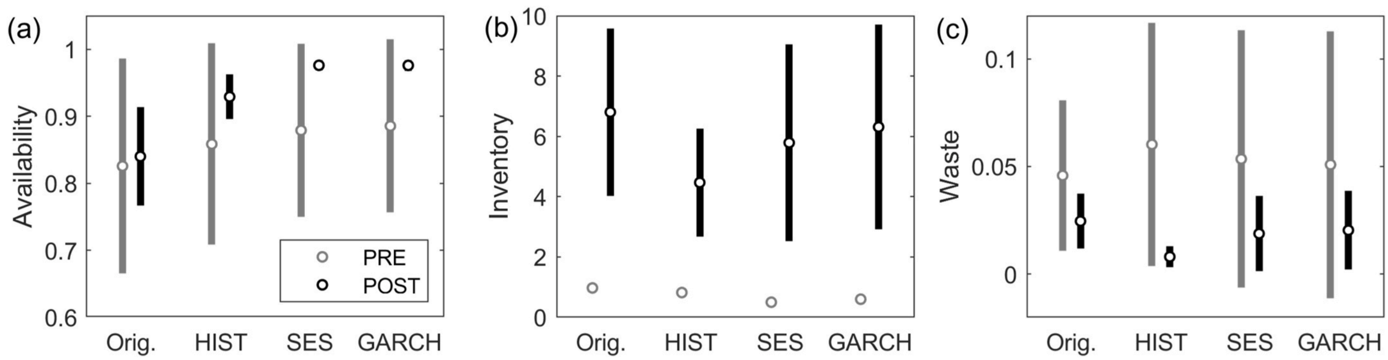

Fig. 5. Comparison of performance measures between the policies (grey: pre-disruption; black: post-disruption; bars: 95% CI).

Table 1
Pair comparison of simulation results under alternative decisions.

		PRE			POST		
		HIST	SES	GARCH	HIST	SES	GARCH
Availability	**Orig.**	0.033 [0.006, 0.059] p = 0.02	0.033 [−0.027, 0.134] p = 0.18	0.060 [−0.018, 0.138] p = 0.12	0.089 [0.003, 0.148] p = 0.005	0.136 [0.065, 0.207] p < 0.001	0.136 [0.066, 0.206] p < 0.001
	HIST	–	0.021 [−0.039, 0.080] p = 0.48	0.027 [−0.030, 0.084] p = 0.33	–	0.047 [0.020, 0.074] p = 0.002	0.047 [0.021, 0.073] p = 0.001
	SES	–	–	0.007 [−0.003, 0.016] p = 0.16	–	–	0.000 [−0.004, 0.004] p = 0.95
Average Inventory	**Orig.**	−0.153 [−0.313, 0.008] p = 0.061	−0.473 [−0.638, −0.309] p < 0.001	−0.373 [−0.544, −0.203] p < 0.001	−2.337 [−4.321, −0.353] p = 0.02	−1.017 [−3.884, 1.851] p = 0.47	−0.494 [−3.431, 2.443] p = 0.73
	HIST	–	−0.321 [−0.450, −0.191] p < 0.001	−0.221 [−0.337, −0.105] p < 0.001	–	1.320 [−0.434, 3.075] p = 0.13	1.843 [−0.018, 3.705] p = 0.05
	SES	–	–	0.100 [0.056, 0.144] p < 0.001	–	–	0.523 [0.306, 0.741] p < 0.001
Average Waste	**Orig.**	0.015 [−0.017, 0.046] p = 0.34	0.008 [−0.027, 0.042] p = 0.64	0.005 [−0.034, 0.044] p = 0.79	−0.017 [−0.029, −0.044] p = 0.009	−0.006 [−0.028, 0.016] p = 0.59	−0.004 [−0.027, 0.018] p = 0.70
	HIST	–	−0.007 [−0.029, 0.015] p = 0.52	−0.009 [−0.034, 0.015] p = 0.43	–	0.011 [−0.007, 0.028] p = 0.21	0.012 [−0.005, 0.030] p = 0.16
	SES	–	–	−0.003 [−0.008, 0.003] p = 0.32	–	–	0.002 [−0.002, 0.006] p = 0.43

rithms can reduce inventory level but not significantly. For the waste generated, the algorithms do not have a significant impact in either the PRE ($F = 0.05$; $p = 0.98$) or the POST ($F = 1.95$; $p = 0.13$). The significance level in the POST period is slightly higher. Between pairs of algorithms, the HIST, SES and GARCH algorithms can reduce waste in the POST period for 67.1%, 52.0% and 17.1% respectively, compared with judgment.

Algorithms (incumbent and alternative) can either increase availability and reduce waste (in the case of POST) or reduce inventory level (in the case of PRE), without harming the inventory performance in the other fronts. Overall, the current algorithm HIST outperforms the judgmental decisions represented by the original order quantities. Further, SES outperforms the judgmental decisions in all three metrics in POST, and its performance is not worse than GARCH. In the PRE period, SES is not worse than judgmental decisions or HIST. Given the computational resources needed in running the GARCH model for estimating the volatility, and the fact that SES requires even less data storage than HIST, we can conclude that SES is an effective and efficient algorithm for safety stock setting during disruptive events.

5. Slack-fitness considerations

We have established that algorithm aversion behaviour is present during disruptions, and it tends to deteriorate inventory management performance. Additionally, algorithms like SES and GARCH can improve performance compared with the incumbent algorithm (HIST). These findings align with previous experimental results on algorithm aversion in volatile environments (Dietvorst and Bharti, 2020). The next important question is why decision-makers drift away from algorithmic recommendations and rely more on subjective judgement. In this section, we aim to generalize our findings and theoretically elucidate algorithm adoption behaviour during disruptions. The proposed slack-fitness theory integrates technology acceptance model (TAM, Davis, 1989), task-technology fit (TTF, Goodhue, 1995; Goodhue and Thompson, 1995) and slack resources theory (Rahrovani and Pinsonneault, 2012). This theoretical framework integrates information systems and behavioural theories and encompasses both individual and organizational levels.

Information systems theories and technology adoption theories—where algorithms can be seen as a special type of decision support technology—can inform our understanding of the algorithm aversion phenomenon. The technology acceptance model (TAM) and task-technology fit (TTF) theories are particularly pertinent to this study. TAM is a popular theory to explain users' acceptance of a technology. Core TAM constructs include *ease-of-use* and *usefulness* of technologies, which influence adoption behaviour (Davis, 1989). On the other hand, TTF focuses on whether technology adoption improves decision performance. It argues that both *task characteristics* and *technology characteristics* determine the *task-technology fit*, which in turn affects *technology use* and *performance benefits* (Goodhue, 1995; Goodhue and Thompson, 1995). Evidently, TAM by itself is insufficient to explain the performance of adopting technology but must be combined with fitness models such as TTF (Dishaw and Strong, 1999; Smith and Mentzer, 2010).

Neither TAM nor TTF can address user behaviour during supply chain disruptions, as they do not contain relevant constructs. Hence, we adapt the theory of slack resources (Rahrovani and Pinsonneault, 2012). The theory of slack resources posits that the value of IT in a business depends on the *IT slack*, defined as extra actual or potential IT resources, which supports IT or organizational adaptation to internal and external pressure. IT slack resources may include time, human resources, and IT artifacts. Slack resources, in general, are defined as unused resources that can be invested. Slack resources are critical for firms to overcome crises (Tognazzo et al., 2016). It can be used to counter threats, explore opportunities, and cope with uncertainty (Cyert and March, 1963; Bourgeois III, 1981; Weinzimmer, 2000). Organizational slack has been identified as a strong antecedent of firm resilience and financial performance in crises (Pal et al., 2014; Li, 2021).

The slack-fitness theory contains the following constructs and factors. We define *disruption* as a severe deviation from the current internal or external operating state. Two kinds of slack are involved: *user slack* and *developer slack*, defined as the slack resources (time) of algorithm users and algorithm developers respectively. A key distinction between the proposed slack-fitness theory and traditional technology adoption theories (TAM and TTF) is a recognition that both algorithm users and developers play parts in algorithm adoption. Asimakopoulos and Dix (2013) recognize the roles of both the designer and the user in forecasting support system adoption, and list lack of training and insufficient knowledge as barriers of adoption in a disruptive event. Users' *perceived algorithm fitness* is defined as users' understanding of how well the algorithm will fit the current circumstance. It is adapted from the usefulness construct in TAM. The other construct in TAM, the ease-of-use property of the algorithm, is unlikely to vary before and after the disruptive event. The *actual algorithm fitness* refers to how well the algorithm will fit in practice. This construct is similar to technology-task fit in TTF. We distinguish between perceived and actual performance to acknowledge users' limited cognitive ability to comprehend the technical details of the algorithm; it is the perceived performance that affects the *adoption*, defined as the rate (or likelihood) of the algorithmic recommendation being implemented in decision-making. Finally, *performance improvement* measures managerial performance (e.g., availability and inventory level) before and after the adoption behaviour.

Fig. 6 shows the theoretical relationships among these various constructs. The plus or minus symbols represent the theorized same-direction or opposite-direction relationships of constructs, respectively. Question marks indicate indeterminate relationships.

A disruption environment negatively affects user adoption behaviour. First, disruptions reduce user and developer slack time, primarily due to increased worker absence and the need to address crises. The effect of increased employee workload is observed during the COVID pandemic (Pamidimukkala and Kermanshachi, 2021), often caused by employee absence due to sickness, quarantine, or family responsibilities. Personnel may also spend more time on firefighting activities during disruptions (Dello Russo et al., 2023).

Reduced slack time, in turn, negatively impacts user familiarity with current algorithms, as comprehensive training requires input from both developers and users. Achieving holistic training is challenging during disruptions. We posit that lack of slack resources during the disruption (e.g., time and infrastructure) hampers algorithm users' ability to understand the algorithms. Similarly, developer constraints impede their capacity to provide user training and develop more suitable algorithms. Rahrovani and Pinsonneault (2015) demonstrated that information systems slack (in terms of technology, knowledge, personnel and time) influences employees' proactive innovation in information technology.

Disruptions may influence adoption by directly diminishing users' belief in the algorithm's task performance. This belief stems not from reduced slack time, but rather from perceived environmental changes resulting from disruptions. Experimental studies have demonstrated that individuals are inclined to reject even superior algorithms if the decision domain and environment are risky and volatile (Dietvorst and Bharti, 2020). The actual performance of algorithms is impacted by the severity of the disruption and developer slack. Developer slack can be utilized to understand the current situation, develop improved algorithms, and update the system for enhanced performance. The disruption alters the environment and the task, thereby influencing the eventual actual performance.

We contend that user adoption behaviour hinges on their perception or comprehension of whether the algorithm fits the current situation. User understanding of the algorithm proves to be a significant precursor to their adoption behaviour. As noted by Yeomans et al. (2019), "It is not enough for [recommendation] systems to be accurate, they must also be

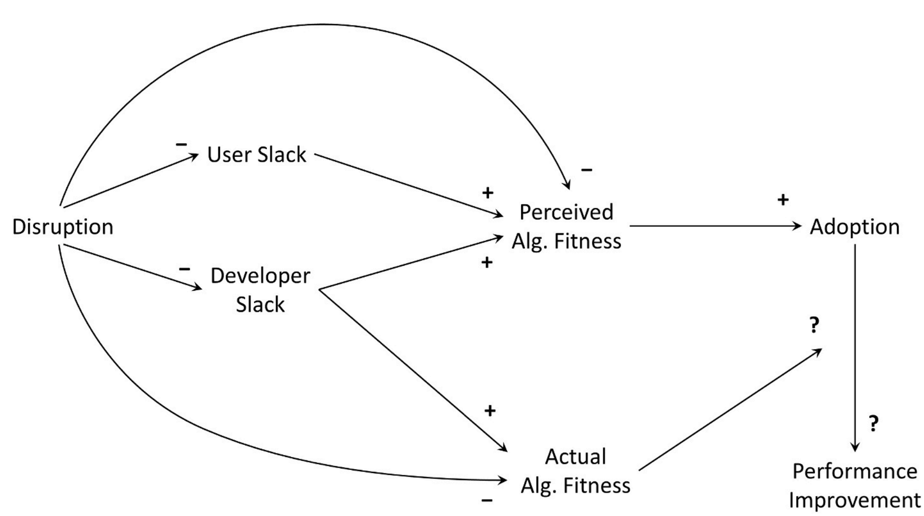

Fig. 6. Algorithm adoption and performance -in-disruption based on Slack-Fitness.

understood." Several factors contribute to this: (1) individuals are more averse to decisions based on complex algorithms than those based on simple algorithms; (2) awareness of algorithm expertise, accuracy, and other pertinent attributes motivate individuals to place greater reliance on algorithms; and (3) individuals are inclined to use algorithms if they are trained in statistical techniques and algorithms (Mahmud et al., 2022). User *lay* theories—based on their understanding and perceptions—regarding how an algorithm functions are also pivotal to their adoption behaviour (Jarupathirun, 2007; Logg et al., 2019).

Furthermore, the actual algorithmic performance plays a crucial role in decision-making. We posit that the actual performance of the algorithm serves as a moderator between adoption and performance, as it determines whether adoption or aversion ultimately proves beneficial. Drawing on TTF theory, we argue that the relations between algorithm aversion and operational performance are contingent on algorithmic fit. This theoretical perspective accommodates counterexamples from published studies that demonstrate algorithm aversion may, at times, improve performance (Fildes et al., 2009).

The slack-fitness theory predicts a reduction in adoption behaviour as the disruptive event unfolds, driven by various mechanisms. However, the resulting performance stemming from this lack of adoption is uncertain and contingent upon the actual fitness of the incumbent algorithm. This theory underscores the significance of user and developer slack as key variables in enhancing algorithmic performance during supply chain disruptions. Moreover, it is worth noting that much of these hypothesized relationships draw support from our initial evidence and insights gleaned from algorithms and decision-making literature. Clearly, further investigation is warranted to validate and refine these theoretical propositions. Consequently, grappling with these complexities can lead to effective decision policies and organizational learning strategies that are adaptable to various types of firms and disruptions.

6. Discussion and conclusions

In this paper, using analytical case study data and simulations akin to a pseudo-natural experiment, we considered the presence and ramifications of algorithm aversion during supply chain disruptions, particularly within the context of safety stock setting. Our findings indicate the following: (1) algorithm aversion changes during supply chain disruptions, lead to disproportionately increased safety stock decisions as risk levels increase; (2) while algorithm aversion detrimentally impacts inventory control performance, simple algorithms demonstrate the potential for significant performance improvement; (3) the propensity for algorithm aversion behaviour during disruptive events can be attributed to information systems slack.

Theoretically, we provide insights into the phenomenon of algorithm aversion. Due to the ex-post nature of our operational data collection, we could not directly measure decision-maker risk attitudes. Therefore, our explanation for algorithm aversion centres on risk estimation rather than risk attitude. Human decision-makers find it more challenging to perceive and predict the changes in the magnitude of uncertainty, measured by second-order metrics such as the variance (Wickens et al., 2020). In our proposed slack-fitness model, we attribute algorithm aversion behaviour to user's diminishing trust in the algorithm and reduced slack of both users and designers during disruptions. The actual fitness of the algorithm acts as a moderator between algorithm adoption (or aversion) behaviour and decision performance. In other words, whether algorithm aversion is beneficial depends on whether the algorithm fits the current situation. However, our study's results imply that users may prematurely disengage from algorithms without fully grasping their suitability for the existing circumstances. This explanation also sheds light on finding by Chae et al. (2014) suggesting that firms with robust IT planning resources (such as mathematical programming, simulation, statistical analysis and machine learning algorithms) tend to achieve higher customer satisfaction.

We have also showcased the effectiveness of pseudo-natural simulation experiment as an assessment tool for algorithms and judgment, complementing traditional laboratory experiments in algorithm aversion research. Given the challenge of directly observing distrust of algorithmic recommendations in practical settings, researchers must focus on revealed effects of trust and distrust, such as the adoption ratios, to measure the extent of algorithm aversion. This approach necessitates in-depth algorithmic understanding and comprehensive environmental variable data to accurately simulate algorithmic recommendations. Additionally, there needs to be clear definitions of decision objectives, such as forecast accuracy or availability. Therefore, this method is well-suited for problems where researchers have access to the necessary information, albeit requiring refinements and adjustments for missing data.

Practically, we propose algorithms that businesses can adopt to more accurately estimate demand risk. It is well-established that firms should increase their safety stock and inventory level during disruptive events due to their mitigating role in hedging against risk (Azadegan et al., 2021; Baghersad and Zobel, 2022). The practical question is how much safety stock should be increased. In this study, we found that the SES and GARCH algorithms demonstrate relatively equal performance in maintaining a high availability level. They outperform both the incumbent algorithm (based on the historical rolling forecast errors) and the judgemental ordering decision in the original dataset. The fact that SES can perform as well as GARCH is particularly promising, given the ease of development and use of SES. Alternatively, advanced forecasting and planning techniques such as GARCH are seldom used in practice (Kanet et al., 2010). The SES smoothing parameter does not significantly affect algorithm performance, further reducing computational complexities. Additionally,Reich et al. (2023) found that if the users are more willing to adopt algorithms if they know the algorithm is adaptive and learning. We suspect that the SES algorithm can enhance user trust due to its highly adaptive nature (SES is otherwise referred to as adaptive smoothing). This observation implies that designers should provide sufficient training to users about the dynamic adjusting ability of the new algorithm, where organizational learning through training is elemental to fit development.

This paper uses a single case study methodology. Several authors argued that the use of single case studies, especially exemplar case studies with unique characteristics, are valuable for scientific development and theory building (Flyvbjerg, 2011; (Yin, 2017)). For the theory proposed herein, this case is an exemplar in terms of: (1) the range of SKUs managed by the case company and its requirement of superior stock management capabilities; (2) the deep penetration of demand forecasting, replenishment, and inventory control techniques and algorithms within the case company; and (3) the COVID-19 pandemic as a representative disruptive event for all supply chain participants. These characteristics increase the generalizability of the proposed theory.

One limitation of this research is the issue of censored demand. In retail contexts, the actual demand is often not directly observable. The observable sales volume is constrained (censored) by the available stock level. Therefore, when the ordering policy is adjusted (e.g., from the incumbent algorithm to SES), the sales volume is likely to change, especially when the stock level is low. In this study, we used the sales data to represent demand across all simulation scenarios. Future research can address the challenge of demand censorship and attempt to estimate the true demand from observable data, thus providing more accurate insights into inventory management strategies during disruptions.

Additionally, the complexities of the slack-fitness theoretical model may require simplification to fully grasp its benefits for organizational policy and decision-making. Currently, the numerous relationships and moderations are only theoretically proposed. While some of these theoretical relationships are demonstrated in our study, others are based on literature and prior research. Given the diversity of environmental contexts—across industries, product types, supply chain structures, and disruption events—further investigation is needed to understand the nuances fully. The pseudo-natural experiment provided by the COVID pandemic disruptions lays an initial foundation for exploring algorithm aversion and its outcomes during crises. These crises may manifest at global, regional, or local levels, but they existence underscores the importance of understanding potential reactions for effective organizational operational planning. It also sets the stage for algorithms of all types including tactical and even strategic planning—and will depend on the scope of disruptions in addition to algorithm and slack considerations.

CRediT authorship contribution statement

Xun Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Vasco Sanchez Rodrigues:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Data curation, Conceptualization. **Emrah Demir:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Joseph Sarkis:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of interest

None.

Acknowledgement

We thank the anonymous referees for their invaluable comments to improve this paper.

Data availability

The authors do not have permission to share data.

References

Alexander, V., Blinder, C., Zak, P.J., 2018. Why trust an algorithm? Performance, cognition, and neurophysiology. Comput. Hum. Behav. 89, 279–288.
Asimakopoulos, S., Dix, A., 2013. Forecasting support systems technologies-in-practice: a model of adoption and use for product forecasting. Int. J. Forecast. 29 (2), 322–336.
Azadegan, A., Modi, S., Lucianetti, L., 2021. Surprising supply chain disruptions: mitigation effects of operational slack and supply redundancy. Int. J. Prod. Econ. 240, 108218.
Baghersad, M., Zobel, C.W., 2022. Organizational resilience to disruption risks: developing metrics and testing effectiveness of operational strategies. Risk Anal. 42 (3), 561–579.
Barros, J., Cortez, P., Carvalho, M.S., 2021. A systematic literature review about dimensioning safety stock under uncertainties and risks in the procurement process. Operations Research Perspectives 8, 100192.
Barto, J., Eruguz, A.S., Spliet, R., Wøhlk, S., 2024. Reduced food waste through inventory control despite throwing out food before expiration: online vs. offline retail. Omega, 103131.
Becker-Peth, M., Thonemann, U.W., Gully, T., 2018. A note on the risk aversion of informed newsvendors. J. Oper. Res. Soc. 69 (7), 1135–1145.
Beutel, A.L., Minner, S., 2012. Safety stock planning under causal demand forecasting. Int. J. Prod. Econ. 140 (2), 637–645.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econom. 31 (3), 307–327.
Boudoukh, J., Richardson, M., Whitelaw, R.F., 1997. Investigation of a class of volatility estimators. J. Deriv. 4 (3), 63–71.
Bourgeois III, L.J., 1981. On the measurement of organizational slack. Acad. Manag. Rev. 6 (1), 29–39.
Bretschneider, S., 1986. Estimating forecast variance with exponential smoothing Some new results. Int. J. Forecast. 2 (3), 349–355.
Browning, T., Kumar, M., Sanders, N., Sodhi, M.S., Thürer, M., Tortorella, G.L., 2023. From supply chain risk to system-wide disruptions: research opportunities in forecasting, risk management and product design. Int. J. Oper. Prod. Manag. 43 (12), 1841–1858.
Cachon, G.P., Lariviere, M.A., 1999. An equilibrium analysis of linear, proportional and uniform allocation of scarce capacity. IIE Trans. 31, 835–849.
Cannella, S., Di Mauro, C., Dominguez, R., Ancarani, A., Schupp, F., 2019. An exploratory study of risk aversion in supply chain dynamics via human experiment and agent-based simulation. Int. J. Prod. Res. 57 (4), 985–999.
Castelo, N., Bos, M.W., Lehmann, D.R., 2019. Task-dependent algorithm aversion. J. Market. Res. 56 (5), 809–825.
Chae, B., Olson, D., Sheu, C., 2014. The impact of supply chain analytics on operational performance: a resource-based view. Int. J. Prod. Res. 52 (16), 4695–4710.
Christopher, M., Holweg, M., 2011. "Supply Chain 2.0": managing supply chains in the era of turbulence. Int. J. Phys. Distrib. Logist. Manag. 41 (1), 63–82.
Corbett, C.J., Fransoo, J.C., 2007. Entrepreneurs and newsvendors: do small businesses follow the newsvendor logic when making inventory decisions? Working paper. Retrieved at. https://escholarship.org/uc/item/2hc5v77z. (Accessed 17 May 2024).
Cyert, R.M., March, J.G., 1963. A behavioral theory of the firm 2 (4), 169–187.
Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 319–340.
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. Eur. J. Oper. Res. 269 (3), 955–983.
Dello Russo, S., Mayrhofer, W., Caetano, A., Passos, A., 2023. High-commitment HRM practices during the financial crisis in Portugal: employees' and HR perspectives. Eur. Manag. Rev. 21 (1), 204–219.
Dietvorst, B.J., Simmons, J.P., Massey, C., 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J. Exp. Psychol. Gen. 144 (1), 114.
Dietvorst, B.J., Simmons, J.P., Massey, C., 2018. Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. Manag. Sci. 64 (3), 1155–1170.
Dietvorst, B.J., Bharti, S., 2020. People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. Psychological science 31 (10), 1302–1314.
Dishaw, M.T., Strong, D.M., 1999. Extending the technology acceptance model with task–technology fit constructs. Information & management 36 (1), 9–21.
Eckel, C.C., Grossman, P.J., 2008. Men, women and risk aversion: experimental evidence. Handbook of experimental economics results 1, 1061–1073.
Eeckhoudt, L., Gollier, C., Schlesinger, H., 1995. The risk-averse (and prudent) newsboy. Manag. Sci. 41 (5), 786–794.
Feng, X., Gao, J., 2020. Is optimal recommendation the best? A laboratory investigation under the newsvendor problem. Decis. Support Syst. 131, 113251.
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. Int. J. Forecast. 25 (1), 3–23.
Fildes, R., Goodwin, P., 2021. Stability in the inefficient use of forecasting systems: a case study in a supply chain company. Int. J. Forecast. 37 (2), 1031–1046.
Filiz, I., Judek, J.R., Lorenz, M., Spiwoks, M., 2023. The extent of algorithm aversion in decision-making situations with varying gravity. PLoS One 18 (2), e0278751.
Flyvbjerg, B., 2011. Case Study. In: Denzin, N.K., Lincoln, Y.S. (Eds.), The Sage Handbook of Qualitative Research, 4th ed. SAGE Publications, Inc, pp. 301–316.
Gonçalves, J.N., Carvalho, M.S., Cortez, P., 2020. Operations research models and methods for safety stock determination: a review. Operations Research Perspectives 7, 100164.
Goodhue, D.L., 1995. Understanding user evaluations of information systems. Manag. Sci. 41 (12), 1827–1844.
Goodhue, D.L., Thompson, R.L., 1995. Task-technology fit and individual performance. MIS Q. 213–236.
Jarupathirun, S., 2007. Dialectic decision support systems: system design and empirical evaluation. Decis. Support Syst. 43 (4), 1553–1570.
Kanet, J.J., Gorman, M.F., Stößlein, M., 2010. Dynamic planned safety stocks in supply networks. Int. J. Prod. Res. 48 (22), 6859–6880.
Kanheman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 363–391.
Katsaliaki, K., Galetsi, P., Kumar, S., 2022. Supply chain disruptions and resilience: a major review and future research agenda. Ann. Oper. Res. 319, 965–1002.
Krupp, J.A., 1997. Safety stock management. Prod. Inventory Manag. J. 38 (3), 11.
Li, Z., 2021. Exploring the role of organizational slack in the COVID-19 pandemic: an empirical study of the manufacturing industry. Corp. Govern.: The International Journal of Business in Society 21 (6), 996–1010.
Logg, J.M., Minson, J.A., Moore, D.A., 2019. Algorithm appreciation: people prefer algorithmic to human judgment. Organ. Behav. Hum. Decis. Process. 151, 90–103.
Longoni, C., Bonezzi, A., Morewedge, C.K., 2019. Resistance to medical artificial intelligence. J. Consum. Res. 46 (4), 629–650.
Lowens, E., 2020. Accuracy is not enough: the task mismatch explanation of algorithm aversion and its policy implications. Harv. J. Law Technol. 34 (1).
MacCrimmon, K.R., Wehrung, D.A., 1990. Characteristics of risk taking executives. Manag. Sci. 36 (4), 422–435.
Mahmud, H., Islam, A.N., Ahmed, S.I., Smolander, K., 2022. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Technol. Forecast. Soc. Change 175, 121390.
MATLAB, 2018. 9.7. 0.1190202 (R2019b). The MathWorks Inc, Natick, Massachusetts.
Morewedge, C.K., 2022. Preference for human, not algorithm aversion. Trends Cognit. Sci. 26 (10), 824–826.
Nahmias, S., 1982. Perishable inventory theory: A review. Operations research 30 (4), 680–708.
Nikookar, E., Yanadori, Y., 2022. Preparing supply chain for the next disruption beyond COVID-19: managerial antecedents of supply chain resilience. International journal of operations & production management 42 (1), 59–90.
Pal, R., Torstensson, H., Mattila, H., 2014. Antecedents of organizational resilience in economic crises—an empirical study of Swedish textile and clothing SMEs. Int. J. Prod. Econ. 147, 410–428.
Pamidimukkala, A., Kermanshachi, S., 2021. Impact of Covid-19 on field and office workforce in construction industry. Project Leadership and Society 2, 100018.

Prahl, A., Van Swol, L., 2017. Understanding algorithm aversion: when is advice from automation discounted? J. Forecast. 36 (6), 691–702.

Rahrovani, Y., Pinsonneault, A., 2012. On the business value of information technology: a theory of slack resources. In: Information Systems Theory. Springer, New York.

Rahrovani, Y., Pinsonneault, A., 2015. Individually perceived is slack resources and innovating with IT. In: 2015 48th Hawaii International Conference on System Sciences. IEEE, pp. 4171–4180.

Reich, T., Kaju, A., Maglio, S.J., 2023. How to overcome algorithm aversion: Learning from mistakes. Journal of Consumer Psychology 33 (2), 285–302.

Rinaldi, M., Murino, T., Gebennini, E., Morea, D., Bottani, E., 2022. A literature review on quantitative models for supply chain risk management: can they be applied to pandemic disruptions? Comput. Ind. Eng., 108329

Scarpin, M.R.S., Scarpin, J.E., Musial, N.T.K., Nakamura, W.T., 2022. The implications of COVID-19: bullwhip and ripple effects in global supply chains. Int. J. Prod. Econ. 251, 108523.

Schleper, M.C., Gold, S., Trautrims, A., Baldock, D., 2021. Pandemic-induced knowledge gaps in operations and supply chain management: COVID-19's impacts on retailing. Int. J. Oper. Prod. Manag. 41 (3), 193–205.

Seuring, S., Brandenburg, M., Sauer, P.C., Schünemann, D.S., Warasthe, R., Aman, S., Qian, C., Petljak, K., Neutzling, D.M., Land, A., Khalid, R.U., 2022. Comparing regions globally: impacts of COVID-19 on supply chains–a Delphi study. Int. J. Oper. Prod. Manag. 42 (8), 1077–1108.

Smith, C.D., Mentzer, J.T., 2010. Forecasting task-technology fit: the influence of individuals, systems and procedures on forecast performance. Int. J. Forecast. 26 (1), 144–161.

Stößlein, M., Kanet, J.J., Gorman, M., Minner, S., 2014. Time-phased safety stocks planning and its financial impacts: empirical evidence based on European econometric data. Int. J. Prod. Econ. 149, 47–55.

Stuart, I., McCutcheon, D., Handfield, R., McLachlin, R., Samson, D., 2002. Effective case research in operations management: a process perspective. J. Oper. Manag. 20 (5), 419–433.

Tognazzo, A., Gubitta, P., Favaron, S.D., 2016. Does slack always affect resilience? A study of quasi-medium-sized Italian firms. Enterpren. Reg. Dev. 28 (9–10), 768–790.

Trapero, J.R., Cardos, M., Kourentzes, N., 2019. Empirical safety stock estimation based on kernel and GARCH models. Omega 84, 199–211.

Weinzimmer, L.G., 2000. A replication and extension of organizational growth determinants. J. Bus. Res. 48 (1), 35–41.

Wickens, C.D., Clegg, B.A., Witt, J.K., Smith, C.A.P., Herdener, N., Spahr, K.S., 2020. Model of variability estimation: factors influencing human prediction and estimation of variability in continuous information. Theor. Issues Ergon. Sci. 21 (2), 220–238.

Yeomans, M., Shah, A., Mullainathan, S., Kleinberg, J., 2019. Making sense of recommendations. J. Behav. Decis. Making 32 (4), 403–414.

Yin, R.K., 2017. Case Study Research and Applications - Design and Methods. SAGE Publications.

Zhang, F., 2007. An application of vector GARCH model in semiconductor demand planning. Eur. J. Oper. Res. 181 (1), 288–297.