INTERNATIONAL JOURNAL OF BUSINESS AND ECONOMICS INSIGHTS

Volume: 15; **Issue:** 8
Pages: 01– 26
Accepted: 17 June 2025
Published: 02 August 2025

Measuring the Hidden Cost of AI Errors in Supply Chain Decision-Making

Tasnim Binte Afsar[1]

[1]. *Master in Business Analytics Management, Avila University, Kansas City Missouri, USA*
Bachelor of Business Administration in Finance, United International University, Dhaka, Bangladesh;
Email: tasnim.afsar3@gmail.com

DOI: 10.63125/y33vkx18
This work was peer-reviewed under the editorial responsibility of the IJBEI, 2025

Abstract

Artificial intelligence (AI) forecasts are judged by their accuracy at the point of decision, but the cost of their errors is paid across the supply chain. This paper asks how large the costs of AI forecast errors are that never appear where the forecast is used, and what determines them. Building on the theory of information distortion, we propose a research model in which AI forecast error raises hidden upstream cost through order distortion and bullwhip amplification, error structure matters at equal accuracy, and human oversight attenuates the effect. We test it in a computational experiment: a three-echelon supply chain in which the retailer plans with an AI forecast whose error structure (white noise, rare spikes, autocorrelation, under- and over-forecast bias), magnitude and oversight regime are manipulated in a full factorial design with 9,000 runs and common random numbers. Errors of identical root mean squared error produced very different costs. Unbiased noise and rare spikes placed 79% and 72% of their cost increase upstream, so that the total cost increase was up to 7.0 times the increase visible at the retailer, whereas over-forecast bias raised only the retailer's cost and under-forecast bias cost little anywhere. Contrary to our second hypothesis, persistent errors generated far less hidden cost than transient ones. Order distortion and upstream bullwhip serially mediated the effect of error magnitude on hidden cost (indirect effect 6.93, 95% CI [6.55, 7.32]). Accuracy measures explained 31% of the variation in the cost increase, against 53% once error structure was also known. A simple override rule cut the hidden cost of rare spikes by 60% but made large biases more costly. A post hoc text analysis of 755 research-article titles shows that research on AI in supply chains and research on the cost of forecast errors rarely meet. The findings imply that AI forecasts should be evaluated by the cost they cause across the chain, not by accuracy at the decision point.

Keywords

Artificial intelligence; forecast error; hidden cost; bullwhip effect; supply chain decision-making; human oversight; simulation experiment; mediation analysis

INTRODUCTION

Firms have moved quickly to place artificial intelligence (AI) at the centre of supply chain planning. Machine learning models now forecast demand, recommend orders and set inventory targets in retail, manufacturing and logistics, and reviews and case studies of the field document a rapid growth of applications across forecasting, procurement, production and risk management (Baryannis et al., 2019; Helo & Hao, 2022; Pournader et al., 2021; Riahi et al., 2021; Toorajipour et al., 2021), building on earlier work on machine learning for supply chain demand forecasting and on big data analytics in operations (Carbonneau et al., 2008; Choi et al., 2018; Syntetos et al., 2016). Digital technologies are also changing how supply chain risks are analysed and how disruptions spread (Dolgui & Ivanov, 2020; Ivanov et al., 2019). The promise is better decisions from better predictions, and the evidence from forecasting competitions shows that machine learning methods can be highly accurate on retail demand (Makridakis et al., 2022). The way these models are evaluated, however, has not changed with their adoption. An AI forecast is almost always judged by an accuracy measure computed where it is used, such as the root mean squared error (RMSE) or the mean absolute error (MAE) of the retailer's demand forecast (Hyndman & Koehler, 2006), and the forecasting literature has long warned that such measures are imperfect guides to the inventory and cost consequences of a forecast (Kourentzes et al., 2020; Syntetos et al., 2010).

The gap matters because a supply chain is a system in which the cost of an error is not paid where the error is made. A forecast error at the retailer changes the retailer's orders, and those orders are the demand signal of the distributor, whose orders in turn drive the manufacturer's production. The amplification of order variability as it moves upstream, the bullwhip effect, has been traced to the way forecasts and replenishment rules process demand information (Chen, Drezner, et al., 2000; Dejonckheere et al., 2003; Lee et al., 1997), and it imposes costs on stages that never see the original forecast. An AI model that is slightly more accurate than its predecessor at the retailer may therefore be more expensive for the chain, and the people who evaluate the model will not see the difference. We call these costs hidden: they are caused by the AI forecast but borne outside the stage and the metric by which the forecast is judged.

Although AI-based technologies can improve supply chain decisions, the cost of their errors has received far less attention than their accuracy. Research on human interaction with algorithms has shown that people lose trust in algorithms after seeing them err (Dietvorst et al., 2015), that they sometimes prefer algorithmic to human judgement (Logg et al., 2019), and that planners frequently adjust statistical forecasts (Fildes et al., 2009; Perera et al., 2019). Field studies have examined what happens when managers override algorithmic recommendations in retail and warehouse operations (Caro & Sáez de Tejada Cuenca, 2023; Kesavan & Kushwaha, 2020; Sun et al., 2022). These studies concern whether people follow the algorithm. The prior question, how much an algorithm's errors cost and where that cost falls, has been addressed for classical forecasting methods (Ali et al., 2012; Fildes & Kingsman, 2011; Sanders & Graman, 2009) but not for the error patterns typical of AI models, which can be unbiased yet unstable, accurate on average yet occasionally grossly wrong, or slow to adapt to change. The research question of this paper is therefore: how large are the hidden costs of AI forecast errors in a supply chain, and how do they depend on the structure of the errors, on the mechanisms that carry them upstream and on human oversight?

This paper addresses the question with a research model grounded in the theory of information distortion and tests it in a controlled computational experiment. In the experiment the retailer of a three-echelon supply chain plans with an AI forecast constructed as a perfect-information forecast plus an injected error whose structure and magnitude are manipulated, while the root mean squared error is held equal across structures by design. This design isolates the question that matters for practice: if two AI models are equally accurate by the usual measure, do their errors cost the same? The contributions are fourfold. First, we define and measure the hidden cost of AI forecast errors as the cost increase borne upstream of the decision point, and we show that it is large and that the accuracy measures used to evaluate AI models explain a minority of it. Second, we show that error structure, not only error size, determines hidden cost, and that the ordering is the reverse of what an emphasis on bias would suggest: unbiased but unstable errors are the most expensive. Third, we identify the mechanism, a serial path from error to order distortion to upstream bullwhip to hidden cost, with bootstrap mediation analysis.

Fourth, we show that a simple human override rule can remove much of the hidden cost of rare large errors but can raise the cost of other error types, so that oversight must be designed for the error structure of the model. A post hoc text analysis of research-article titles complements the experiment by showing how rarely research on AI in supply chains addresses the cost of AI errors.

The rest of the paper is organised as follows. The next section develops the theoretical support and the hypotheses. The methodology section describes the simulation model, the experimental design and the measures. The results section reports the manipulation check, the hypothesis tests, the mediation and moderation analyses, robustness checks and the post hoc text analysis. The final section discusses the theoretical and managerial implications, the limitations and directions for future research.

THEORETICAL SUPPORT AND RESEARCH MODEL

The theory of information distortion in supply chains explains why decisions taken on locally processed information create costs elsewhere in the chain (Lee et al., 1997). Its central claim is that each stage transforms the demand information it receives through its forecasting and replenishment rules before passing it upstream as orders, and that this transformation amplifies variability. The causes identified in the original formulation include demand signal processing, order batching, price fluctuation and rationing (Lee et al., 1997); subsequent work quantified the amplification produced by forecasting with moving averages and exponential smoothing under order-up-to policies (Chen, Drezner, et al., 2000; Chen, Ryan, et al., 2000), showed with control theory that such policies amplify variability for common forecasting methods (Dejonckheere et al., 2003; Disney & Towill, 2003), and examined how demand that follows an autoregressive process evolves as it moves through the chain (Zhang, 2004). Behavioural experiments showed that decision makers misperceive the feedback in these systems and amplify variability further (Croson & Donohue, 2006; Sterman, 1989), and empirical studies found that the strength of the effect varies across industries and with the information transmitted between stages (Bray & Mendelson, 2012; Cachon et al., 2007), as a review of the field documents (Wang & Disney, 2016).

From this lens, an AI demand forecast is one more processor of demand information, and its errors are a new source of distortion. Two features distinguish this view from the usual evaluation of forecasts. First, the effect of a forecast error is mediated by the replenishment rule: under an order-up-to policy the order responds to the change in the forecast, multiplied by the length of the protection interval, rather than to the forecast error itself. Second, the cost of the distorted orders is borne upstream, by stages whose costs are not part of the forecast's evaluation. Research on forecast evaluation for inventory decisions has made the related point that accuracy measures and inventory performance can diverge (Kourentzes et al., 2020; Syntetos et al., 2010) and that safety stocks must reflect the uncertainty of the forecast (Prak et al., 2017), and research on decision-focused learning argues that predictive models should be trained and evaluated by the decisions they induce rather than by their accuracy (Ban & Rudin, 2019; Bertsimas & Kallus, 2020; Elmachtoub & Grigas, 2022), a refinement of the broader distinction between explanatory and predictive modelling (Shmueli & Koppius, 2011). Our model extends this reasoning from the decision at hand to the decisions it triggers along the chain.

A second theoretical foundation is the economics of supply chain coordination, which explains why the costs created at one stage are rarely corrected by the stage that causes them (Chowdary & Md, 2020; Chowdary & Jamil, 2020). When firms in a chain make decisions independently, each one optimises its own objective and ignores the effect of its choices on its partners (Asheke & Bin, 2022; Mohsin & Jannatul, 2021; Chowdary, 2021). The result is the familiar problems of double marginalisation, misaligned incentives and system-wide outcomes that fall short of what a centralised decision maker would achieve. Contracting research has shown that buyback, revenue-sharing, quantity-flexibility and capacity-reservation agreements can realign incentives so that a downstream party bears part of the cost its decisions impose upstream (Islam & Md, 2023; Khalid & Sadmeen, 2022; Maksudul & Tazul Islam, 2022). Most of these mechanisms were designed around quantities and prices, however, not around the quality of the information that drives orders (Mashfiqur &Jamil, 2023; Md & Kynatun, 2023; Maksudul & Hossen, 2023). Forecasting choices are generally treated as an internal matter for the firm that makes them. The upstream cost of a volatile forecast therefore acts as an externality: the buyer gains the accuracy benefits of a new model, while the supplier absorbs the variability it creates (Mohsin & Kaniz, 2023; Mamunur & Md, 2023; Sadia & Mosheur, 2023). Without a contractual or organisational

mechanism that prices this variability, there is little reason to expect firms to account for it when they select and tune forecasting models (Sharmin & Kaniz, 2023; Sharmin & Nishat, 2023; Sadmeen & Khalid, 2023).

Organisational information processing theory offers a complementary view of why AI forecasts change the way uncertainty moves through the chain. The theory holds that organisations must match their information processing capacity to the uncertainty they face (Gazi Yeasin, 2024a, 2024b; Maishar Mostari, 2024). They can do this by building buffers and slack that reduce the need for information, or by investing in systems and lateral relationships that increase their ability to process it. AI forecasting is a clear example of the second route (Mashfiqur, 2024; Masrul & Mashfiqur, 2024; Masrul & Kanti, 2024). It greatly increases a firm's capacity to absorb signals from point-of-sale data, promotions, weather, prices and other external sources. Yet greater processing capacity at one stage does not translate into less uncertainty for the chain as a whole (Hasinur, 2024; Maksudul, 2024; Mohsin, 2024). If the faster processing produces more frequent and larger changes in what is passed upstream, the downstream firm reduces its own uncertainty by exporting it to partners whose processing capacity and buffers have not changed (Islam, 2024; Mainuddin, 2024; Mamunur, 2024). This asymmetry suggests that the benefits of AI forecasting depend not only on model quality but on how well the information processing capacities of connected firms match one another.

A third strand concerns the economic value of forecasts and the loss functions used to judge them. Research on forecast evaluation has long noted that statistical accuracy and economic value are different properties. A forecast's value depends on the decision it informs, the cost structure of that decision and the asymmetry between the costs of over- and under-prediction (Songeta, 2024; Sumaiya, 2024; Tahmedur, 2024). In inventory settings, newsvendor-type models show that the optimal decision depends on the whole forecast distribution and on the ratio of shortage to holding costs. Symmetric error measures that treat all deviations alike can therefore favour models that perform poorly in practice. Extending this argument along the supply chain, the relevant loss function is not only that of the focal firm but the combined loss across the stages its orders affect. A forecast that minimises local error may be poorly suited to the chain if its errors take a form that upstream cost structures penalise heavily, such as sudden revisions, persistent bias or synchronised swings. Taken together, these three perspectives provide the basis for the research model. Coordination theory explains why upstream costs go unpriced, information processing theory explains how AI forecasting shifts uncertainty upstream, and forecast value theory explains why local accuracy is an incomplete guide to chain-wide performance.

AI forecast error and hidden cost

Supply chain applications of AI are evaluated where they are deployed. A demand forecasting model at a retailer is assessed on the retailer's data and, at best, on the retailer's inventory performance (Fildes et al., 2022; Seaman, 2018). If its errors distort the retailer's orders, however, the distributor and the manufacturer face more variable demand, carry more safety stock, incur more backorders and change production more often. These costs are real but are recorded in other functions or other firms. Information distortion theory therefore predicts that the cost of AI forecast errors includes a component outside the decision point, and that this component grows with the magnitude of the error, because larger forecast changes produce larger order changes.

H1. AI forecast error increases the hidden cost of the supply chain, borne upstream of the decision point, and the increase grows with the magnitude of the error.

Error structure at equal accuracy

Accuracy measures summarise an error series by one number, but errors with the same root mean squared error can differ in structure. Classical forecast evaluation pays particular attention to bias, because a persistent error accumulates in inventory or backorders, and studies have quantified how forecast errors raise planning and operating costs (Fildes & Kingsman, 2011; Sanders & Graman, 2009); research on judgmental forecasting has documented systematic biases in human forecasts and adjustments (Arvan et al., 2019; Lawrence et al., 2006). AI models can add other structures: errors that are unbiased but unstable from period to period, rare gross errors when the input data are unusual, and slowly drifting errors after a change the model has not learned. If bias is the dominant defect, persistent errors should be the most costly also for the chain, because a persistent error keeps the

retailer's orders away from demand for long periods and propagates as a sustained shift in the orders upstream.

H2. At equal forecast accuracy, persistent AI forecast errors (bias and autocorrelated errors) increase hidden cost more than transient errors (white noise and rare spikes).

Order distortion and upstream amplification

The mechanism through which a forecast error reaches upstream stages is the order. Under an order-up-to policy, the retailer's order equals the demand just observed plus the change in the order-up-to level, and the change in the order-up-to level contains the change in the forecast multiplied by the protection interval (Chen, Drezner, et al., 2000; Dejonckheere et al., 2003). A forecast error therefore first distorts the retailer's orders relative to the orders it would place with a perfect forecast. The distributor and the manufacturer then forecast and replenish from these distorted orders, so the distortion is amplified at each stage, the bullwhip effect (Lee et al., 1997; Wang & Disney, 2016). Finally, the amplified variability raises upstream safety stocks, backorders and production changes, which constitute the hidden cost. The theory thus implies a serial chain of effects.

H3. Order distortion at the decision point and upstream bullwhip amplification serially mediate the effect of AI forecast error on hidden cost.

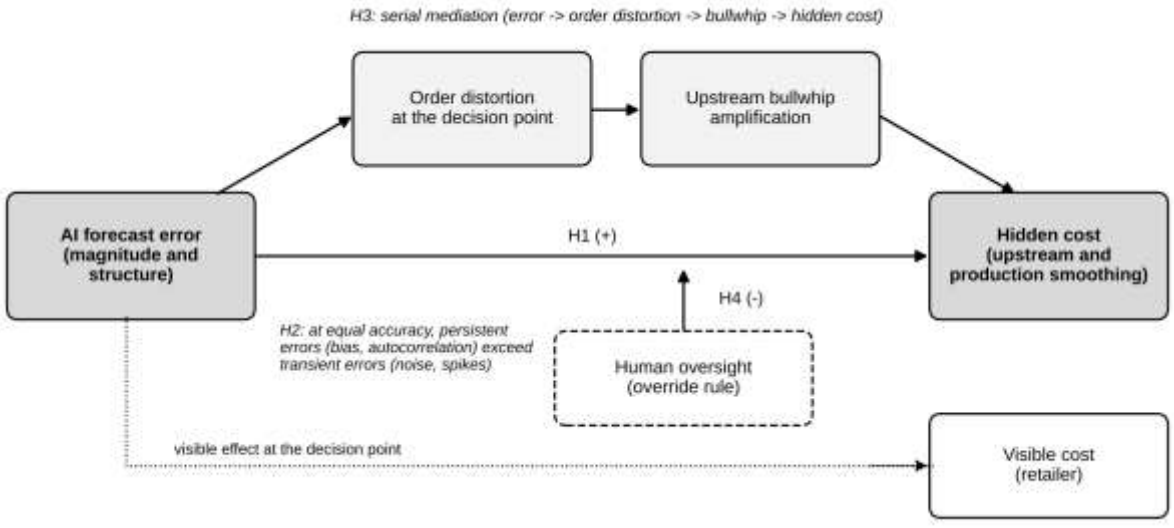
Human oversight

Organisations rarely let AI forecasts drive orders without any human review. Planners adjust statistical forecasts, often when the forecast departs markedly from their own expectation (Brau et al., 2023; Fildes et al., 2009; Perera et al., 2019), and research on human-AI collaboration discusses how to combine algorithmic and human judgement, how trust in AI forms and how the explainability of models and the structure of organisational decisions shape that combination (Burton et al., 2020; Glikson & Woolley, 2020; Khosrowabadi et al., 2022; Rai, 2020; Raisch & Krakowski, 2021; Shrestha et al., 2019). A common rule of thumb is to review forecasts that deviate strongly from a simple benchmark and to blend them with it. Such a rule is most likely to catch errors that are large and sudden, because these produce large deviations, and less likely to catch persistent errors, which a smoothing benchmark partly follows, or small unstable errors, which stay within the tolerance. Oversight should therefore attenuate the hidden cost of AI errors, and the attenuation should be strongest for rare large errors.

H4. Human oversight of AI forecasts attenuates the hidden cost of AI forecast errors, and the attenuation is strongest for rare large errors.

Figure 1 shows the research model with the four hypotheses. The model also distinguishes the visible cost borne at the decision point, which is the cost that an evaluation of the AI forecast at the retailer would record.

Figure 1: Research model: AI forecast error, order distortion, upstream bullwhip amplification, hidden cost and human oversight



METHODOLOGY

The hypotheses concern causal effects of error properties that cannot be manipulated in a field setting, because the error structure of a deployed model is not under the researcher's control and the costs of upstream partners are rarely observable. We therefore test them in a computational experiment, a design widely used in research on the bullwhip effect and on forecasting for inventory control (Chen, Drezner, et al., 2000; Dejonckheere et al., 2003; Kourentzes et al., 2020), and follow established guidance on the design and analysis of simulation experiments (Kleijnen, 2005). The experiment replaces the human participants of a behavioural experiment (Sadeghi R. et al., 2024) with replications of a simulated supply chain: each replication is a five-year history of weekly operations, and the treatments are assigned to the AI forecast rather than to people.

Simulation model

Figure 2 and Table 1 describe the model. A serial supply chain links a retailer, a distributor and a manufacturer, which is supplied from an unlimited source. Every stage reviews its position weekly and places an order; orders placed at the end of a week arrive at the start of the week two periods later when the upstream stage has stock, and are backlogged otherwise. Customer demand (truncated at zero, which is negligible at these parameters) has a mean of 100 units per week with first-order autoregressive innovations (phi = 0.5), a form commonly used in bullwhip research (Zhang, 2004), in three regimes: stationary (innovation standard deviation 10), seasonal (an annual cycle of plus or minus 20%) and volatile (standard deviation 20). Each stage follows an order-up-to policy whose level is the forecast of demand over the protection interval plus a safety stock computed from the root mean square of its own realised lead-time forecast errors over the last 26 weeks, an empirical approach recommended when demand is forecast (Prak et al., 2017; Trapero et al., 2019). The distributor and the manufacturer forecast their incoming orders by exponential smoothing (alpha = 0.3), the standard assumption of the bullwhip literature (Chen, Ryan, et al., 2000). Holding costs rise downstream (0.4, 0.6 and 1.0 per unit and week), backorders cost 10 per unit and week at the retailer and 3 at the internal stages, and the manufacturer pays 0.5 per unit of change in its production order, a capacity adjustment cost that captures the operational burden of variability.

Figure 2: Architecture of the simulation model: three-echelon serial supply chain, AI forecast at the retailer, oversight gate and cost accounting

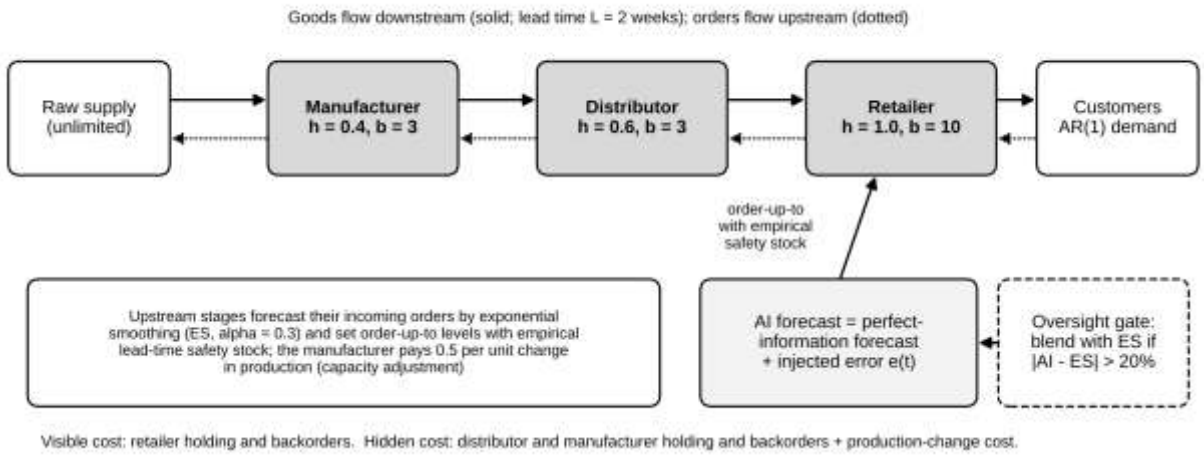


Table 1: Parameters of the simulation model

Parameter	Value	Rationale
Echelons	Retailer, distributor, manufacturer; unlimited raw supply	Serial chain used in bullwhip studies (Chen, Drezner, et al., 2000; Lee et al., 1997)
Review and lead time	Weekly review; L = 2 weeks at every stage	Orders placed at the end of week t arrive at the start of week t + L
Horizon	52-week warm-up, 260 measured weeks	Five years of weekly operation per replication
Customer demand	Mean 100; AR(1) innovations with phi = 0.5; three regimes	Autocorrelated demand as in (Zhang, 2004)
Ordering policy	Order-up-to with empirical lead-time safety stock	Safety stock from realised lead-time errors (Prak et al., 2017; Trapero et al., 2019)
Upstream forecasting	Exponential smoothing of incoming orders, alpha = 0.3	Standard practice and the bullwhip literature (Chen, Ryan, et al., 2000)
Holding cost per unit-week	1.0 (retailer), 0.6 (distributor), 0.4 (manufacturer)	Value added downstream
Backorder cost per unit-week	10 (retailer), 3 (distributor and manufacturer)	Customer service valued above internal delay
Production-change cost	0.5 per unit change in the manufacturer's order	Capacity adjustment cost of order variability
Safety factors	Critical fractile times 1.2 (retailer) and 1.5 (upstream)	Grid search: the perfect-forecast benchmark approximately minimises chain cost
Oversight rule	Blend AI and ES forecasts equally when they differ by more than 20%	Judgmental adjustment heuristic (Fildes et al., 2009)
Replications	100 per cell; common random numbers	Variance reduction for paired comparisons (Kleijnen, 2005)

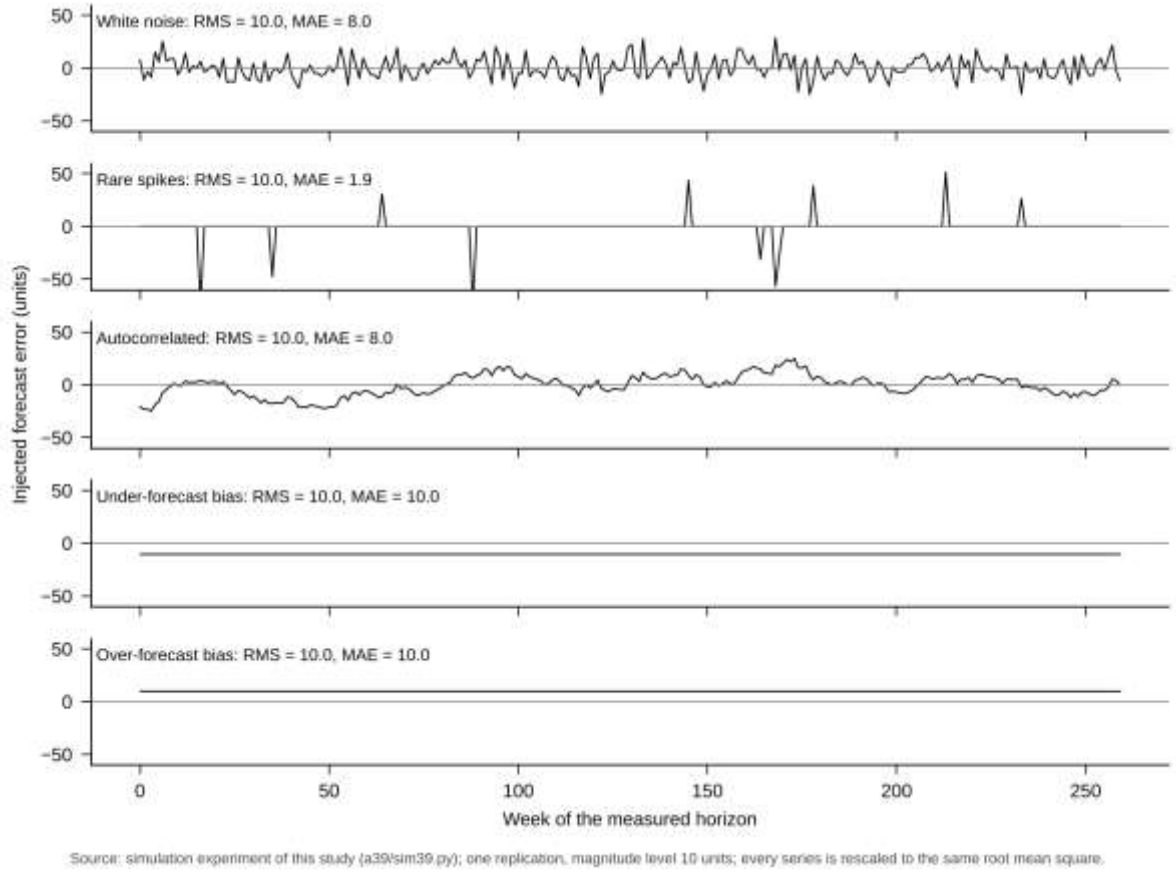
Experimental design and manipulations

The retailer plans with an AI forecast constructed as the perfect-information forecast, the conditional expectation of demand given the true demand process, plus an injected error. Constructing the forecast in this way separates the properties of the error from those of any particular learning algorithm and allows the error to be manipulated exactly. Five error structures were manipulated (Table 2): white noise, rare spikes (an error on about 3% of weeks, large and of random sign, with at least two spikes in every replication), an autocorrelated error with persistence 0.95, and constant under- and over-forecast biases. Each structure was crossed with three magnitudes, defined as the root mean square of the injected error over the measured horizon (5, 10 and 20 units, that is, half, once and twice the standard deviation of the demand innovation in the stationary regime). Every error series was rescaled to its target root mean square in every replication, so that within a magnitude level all structures have the same accuracy as measured by RMSE (Figure 3). The third factor was human oversight: under full automation the AI forecast was used as is; under oversight a planner compared it with an exponential-smoothing forecast of demand and, when the two differed by more than 20%, used their average, a stylised version of the judgmental adjustment observed in practice (Fildes et al., 2009). The fourth factor was the demand regime.

Table 2: Manipulated structures of the AI forecast error

Error structure	Generating process	Persistence	AI failure mode it represents
White noise	Independent normal error each week	None	Unstable predictions from an over-fitted or noisy model
Rare spikes	Error on about 3% of weeks, large and of random sign	None	Occasional gross errors, for example on promotions, data glitches or unseen inputs
Autocorrelated	AR(1) error with rho = 0.95	High	Slow drift after a change the model has not learned
Under-forecast bias	Constant negative error	Complete	Systematic under-prediction, for example censored training data
Over-forecast bias	Constant positive error	Complete	Systematic over-prediction, for example optimistic sales inputs

Figure 3: The five injected error structures in one illustrative draw at the same root mean square (10 units)



Source: simulation experiment of this study (a39/sim39.py); one replication, magnitude level 10 units; every series is rescaled to the same root mean square.

Table 3 summarises the full factorial design: five structures, three magnitudes, two oversight conditions and three regimes, with 100 replications per cell, giving 9,000 experimental runs and 600 benchmark runs with the perfect forecast. Common random numbers were used: within a regime, replication k of every condition faces the same demand path and the same standardised error innovations, and the benchmark run of replication k under the same oversight condition provides its counterfactual. Each run covers a 52-week warm-up and 260 measured weeks.

Table 3: Experimental design: factors, levels and number of runs

Factor	Levels	Number
Error structure	White noise, rare spikes, autocorrelated, under-forecast bias, over-forecast bias	5
Error magnitude (RMS)	5, 10, 20 units	3
Human oversight	Full automation, override rule	2
Demand regime	Stationary, seasonal, volatile	3
Replications per cell	Independent demand paths, shared across conditions (common random numbers)	100
Experimental runs	90 cells times 100 replications	9,000
Benchmark runs	Perfect forecast, by oversight condition and regime	600

Measures

Table 4 defines the variables. The independent variables are the magnitude of the AI error (its root mean square, expressed per 10 units in the regressions) and its structure. The first mediator, order distortion, is the standard deviation of the difference between the retailer's orders and the orders it placed in the paired benchmark run; it measures how much of the forecast error the replenishment rule passes on. The second mediator, upstream bullwhip, is the ratio of the variance of the manufacturer's orders to the variance of customer demand (Chen, Drezner, et al., 2000; Wang & Disney, 2016). The outcomes are cost changes relative to the paired benchmark run, per week: the visible cost change is the change in the retailer's holding and backorder cost, which an evaluation of the AI forecast at the retailer would record; the hidden cost change is the change in the distributor's and the manufacturer's holding and backorder costs plus the production-change cost; their sum is the total cost change. Forecast accuracy was recorded as the RMSE, MAE and mean absolute percentage error of the forecast actually used against realised demand (Hyndman & Koehler, 2006). The demand regime serves as a control.

Table 4: Variables, roles and definitions

Variable	Role	Definition
AI error magnitude (X)	Independent	Root mean square of the injected error: 5, 10 or 20 units (0.5, 1 or 2 times the demand innovation s.d. of 10)
Error structure	Independent	One of five structures (Table 2), matched in root mean square within each magnitude
Human oversight	Moderator	Full automation or the override rule of Table 1
Order distortion (M1)	Mediator	Standard deviation of the difference between the retailer's orders and its orders under the perfect forecast
Upstream bullwhip (M2)	Mediator	Variance of the manufacturer's orders divided by the variance of customer demand
Visible cost change	Outcome	Change in the retailer's holding and backorder cost per week, relative to the benchmark
Hidden cost change (Y)	Outcome	Change in distributor and manufacturer holding and backorder cost plus production-change cost per week
Total cost change	Outcome	Visible plus hidden cost change
Demand regime	Control	Stationary (s.d. 10), seasonal (amplitude 20%) or volatile (s.d. 20)

Benchmark calibration

The safety factors were calibrated so that the perfect-forecast benchmark is close to cost-minimising: a grid search over multipliers of the critical-fractile safety factor (1.0 to 2.0 at the retailer, 1.0 to 2.5 upstream) selected 1.2 at the retailer and 1.5 upstream, which minimised the mean chain cost of the

benchmark across regimes on this grid (at the upstream stages the critical fractile of the distributor's cost ratio is used for both stages). Without this step a biased forecast could appear to reduce cost merely by compensating for a mis-set service level. Table 5 reports the benchmark. Even with a perfect forecast at the retailer, 71% of the chain's cost is incurred upstream and the manufacturer's orders are 12.2 times as variable as customer demand, reflecting that two of the three stages and the production-change cost lie upstream and that the upstream stages forecast from orders by exponential smoothing; the experimental effects below are measured against this benchmark.

Table 5: Performance of the perfect-forecast benchmark by demand regime (full automation)

Regime	Total cost	Retailer cost	Upstream cost	Upstream share	Retail fill rate	Bullwhip ratio (M)	Forecast RMSE
Stationary	128.0	37.1	90.9	71%	0.994	16.1	10.0
Seasonal	132.9	37.6	95.3	72%	0.993	7.3	10.0
Volatile	246.9	73.4	173.5	70%	0.988	13.0	20.0

Analysis

The analysis follows the sequence of the example design in decision support research (Sadeghi R. et al., 2024): a manipulation check, descriptive statistics and correlations, regression models for the hypotheses, a mediation analysis and a moderation analysis. H1 was tested with a one-sided t test of the mean hidden cost change and with regressions of the hidden cost change on error magnitude; H2 with a comparison of persistent and transient structures at equal magnitude; H3 with a serial mediation model (two mediators in sequence) estimated by ordinary least squares with percentile bootstrap confidence intervals from 5,000 resamples, which avoids the limitations of the causal-steps approach (Baron & Kenny, 1986; Hayes, 2009; Preacher & Hayes, 2008); and H4 with paired comparisons of the oversight conditions and with structure-by-oversight interactions. Regression standard errors are heteroskedasticity-consistent (HC3). Because the number of runs is large, almost every effect is statistically significant, so the interpretation rests on effect sizes and confidence intervals. The model, the experiment and every analysis are implemented in Python scripts supplied with the article.

RESULTS

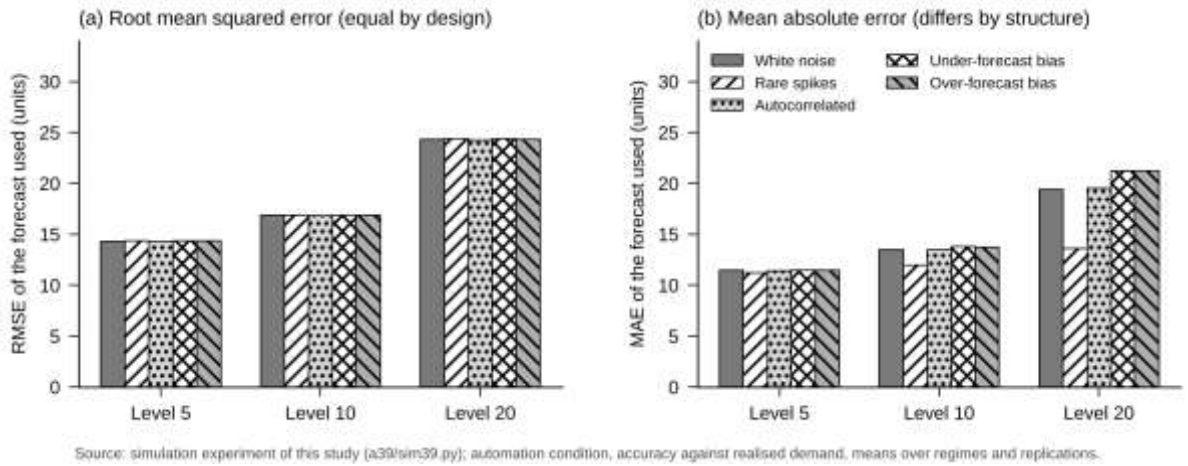
Manipulation check

The manipulation required that error structures be equally accurate by the standard measure within each magnitude level while differing in structure. Table 6 and Figure 4 confirm it. The RMSE of the forecast actually used against realised demand did not differ across structures at any level (level 5: $F = 0.00$, $p = 1.000$; level 10: $F = 0.02$, $p = 1.000$; level 20: $F = 0.07$, $p = 0.990$). The MAE, by contrast, differed at levels 10 and 20 ($F = 16.3$ and 495.6, both $p < 0.001$): at the highest level, rare spikes had an MAE of 13.6 units against 19.4 for white noise and 21.2 for over-forecast bias, so an evaluation by MAE would rank the spiky forecast as the most accurate. Accuracy alone, in other words, does not identify the error structure, and different accuracy measures can rank the same forecasts differently (Gneiting, 2011; Hyndman & Koehler, 2006).

Table 6: Manipulation check: accuracy of the forecast used, by error structure and magnitude (full automation)

Accuracy measure	Noise	Spikes	Autocorr.	Under-bias	Over-bias	One-way ANOVA
Level 5: RMSE	14.32	14.33	14.30	14.33	14.33	$F = 0.00$, $p = 1.000$
Level 5: MAE	11.43	11.18	11.42	11.48	11.46	$F = 0.3$, $p = 0.860$
Level 10: RMSE	16.87	16.90	16.83	16.89	16.88	$F = 0.02$, $p = 1.000$
Level 10: MAE	13.46	11.97	13.46	13.77	13.74	$F = 16.3$, $p < 0.001$
Level 20: RMSE	24.31	24.37	24.25	24.34	24.33	$F = 0.07$, $p = 0.990$
Level 20: MAE	19.39	13.59	19.51	21.25	21.22	$F = 495.6$, $p < 0.001$

Figure 4: Manipulation check: RMSE equal by design (a) and MAE differing by error structure (b)



Source: simulation experiment of this study (a39/sim39.py); automation condition, accuracy against realised demand, means over regimes and replications.

Descriptive statistics and correlations

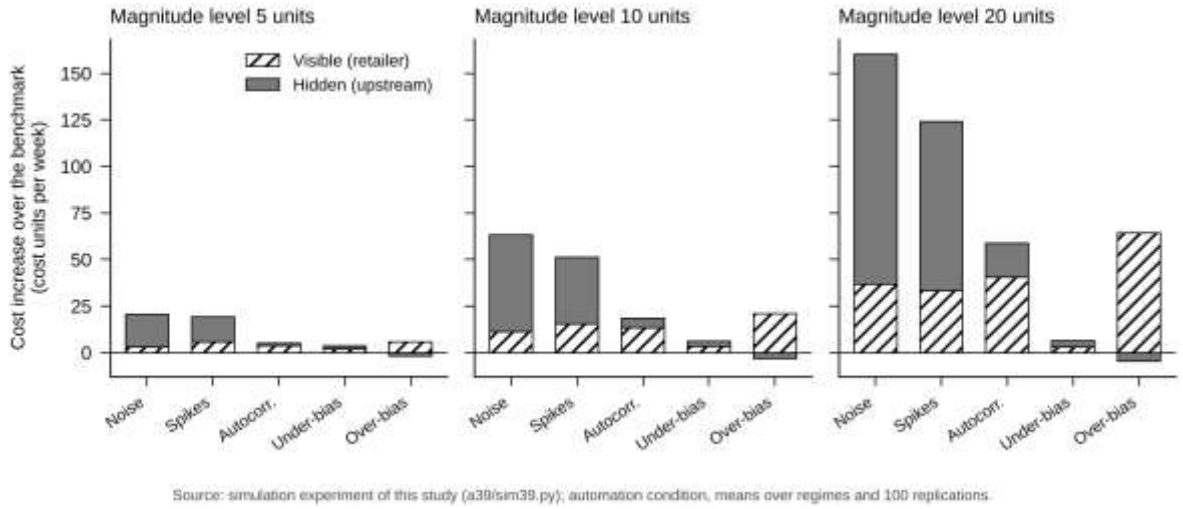
Table 7 reports the cost changes under full automation. Every structure raised the total cost of the chain above the benchmark at every magnitude, but the size and the location of the increase differed sharply (Figure 5). At a magnitude of 10 units, white noise raised total cost by 63.3 per week (37.4% of the benchmark), of which only 11.3 was visible at the retailer, and rare spikes raised it by 51.1, of which 15.2 was visible. An over-forecast bias of the same accuracy raised total cost by 17.6: the retailer's cost rose by 20.7, where the excess stock is held, while the upstream stages saw slightly lower cost. Under-forecast bias was the least costly structure (6.0), most likely because the empirical safety stock, which is computed from realised errors including their bias, rose to absorb most of the shortfall. The ratio of the total to the visible cost increase, the factor by which an evaluation at the retailer understates the cost of the error, ranged from 4.4 to 7.0 for white noise and from 3.3 to 3.7 for rare spikes, against 1.34 to 1.45 for autocorrelated errors.

Table 7: Cost changes relative to the benchmark by error structure and magnitude (full automation, cost units per week)

Error structure	RMS	Total change, mean (s.d.)	Visible change	Hidden change, mean (s.d.)	Hidden share	Total / visible	Total change, % of benchmark
White noise	5	20.5 (9.4)	2.9	17.5 (7.3)	86%	7.02	12.1%
White noise	10	63.3 (20.3)	11.3	51.9 (15.7)	82%	5.59	37.4%
White noise	20	160.2 (34.1)	36.8	123.4 (25.4)	77%	4.36	94.6%
Rare spikes	5	19.2 (10.6)	5.5	13.6 (6.4)	71%	3.47	11.3%
Rare spikes	10	51.1 (19.9)	15.2	35.8 (13.1)	70%	3.35	30.2%
Rare spikes	20	124.1 (35.7)	33.3	90.8 (29.0)	73%	3.73	73.3%
Autocorrelated	5	5.0 (4.0)	3.7	1.3 (1.8)	25%	1.34	3.0%
Autocorrelated	10	18.3 (8.6)	13.3	5.0 (4.1)	27%	1.38	10.8%
Autocorrelated	20	58.8 (19.2)	40.5	18.3 (10.2)	31%	1.45	34.7%
Under-forecast bias	5	3.6 (2.0)	1.8	1.7 (0.4)	49%	1.94	2.1%
Under-forecast bias	10	6.0 (2.9)	3.3	2.8 (0.7)	46%	1.84	3.6%
Under-forecast bias	20	6.4 (4.0)	3.0	3.5 (1.3)	54%	2.17	3.8%
Over-forecast bias	5	3.7 (3.5)	5.6	-1.9 (0.4)	n/a	0.66	2.2%

Error structure	RMS	Total change, mean (s.d.)	Visible change	Hidden change, mean (s.d.)	Hidden share	Total / visible	Total change, % of benchmark
Over-forecast bias	10	17.6 (7.7)	20.7	-3.2 (0.7)	n/a	0.85	10.4%
Over-forecast bias	20	59.6 (13.5)	64.1	-4.5 (1.5)	n/a	0.93	35.2%

Figure 5: Decomposition of the cost increase into its visible (retailer) and hidden (upstream) parts by error structure and magnitude



Source: simulation experiment of this study (a39/sim39.py); automation condition, means over regimes and 100 replications.

Table 8 gives the correlations. The hidden cost change correlated strongly with order distortion (r = 0.95) and with upstream bullwhip (r = 0.80), but weakly with the RMSE of the forecast (r = 0.24) and hardly at all with its MAE (r = 0.03). Regressing the total cost change on RMSE and MAE alone explained 31% of its variance; adding the error structure and the regime raised the share to 81%. Across the 45 cells of the automation condition, the rank correlation between MAE and total cost change was 0.19.

Table 8: Means, standard deviations and correlations of the main variables (full automation, N = 4,500 runs)

Variable	Mean	s.d.	1	2	3	4	5	6	7	8
1. Error RMS	11.67	6.24	1							
2. RMSE	18.51	5.72	0.74	1						
3. MAE	14.55	4.80	0.67	0.93	1					
4. Order distortion	13.45	15.34	0.40	0.29	0.09	1				
5. Bullwhip (M)	20.27	18.61	0.34	0.16	-0.02	0.71	1			
6. Visible change	17.40	19.22	0.69	0.42	0.37	0.29	0.29	1		
7. Hidden change	23.75	38.34	0.44	0.24	0.03	0.95	0.80	0.34	1	
8. Total change	41.15	48.32	0.62	0.36	0.17	0.87	0.75	0.66	0.93	1

Hypotheses testing

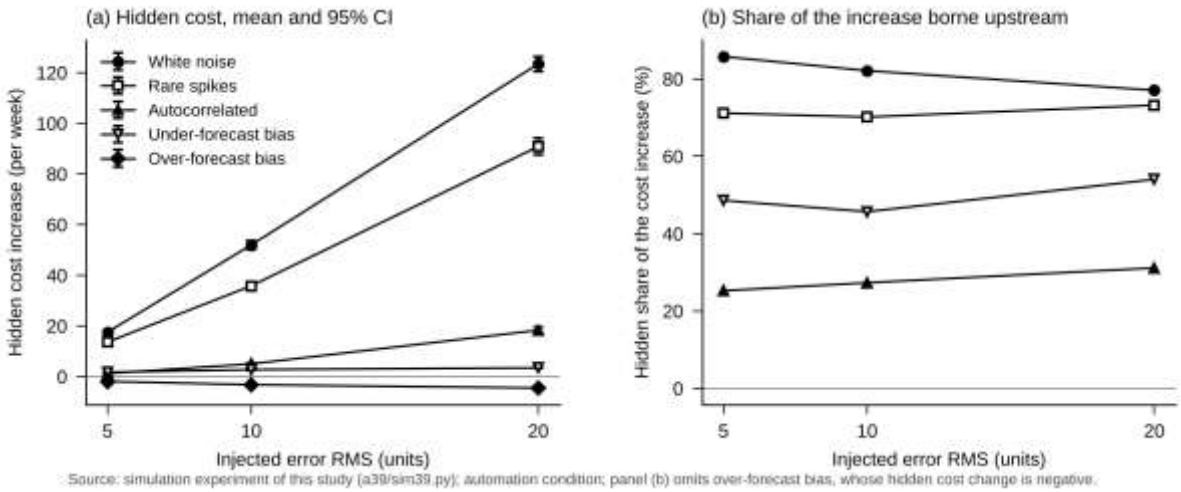
Table 9 reports four regression models of the hidden cost change, following the sequence of control, main effect, structure and moderation models. Model 1 contains the regime controls. Model 2 adds the error magnitude: each additional 10 units of root mean square raised hidden cost by 19.62 per week (SE 0.59, $p < 0.001$), and the explained variance rose from 0.026 to 0.177. The mean hidden cost change

under automation was 23.7 per week (t = 41.5, p < 0.001, one-sided), and it was positive in 77% of the runs. Hidden cost was positive and rose with magnitude for four of the five structures (Figure 6); the exception was over-forecast bias, whose hidden cost change was slightly negative (-3.2 per week on average) most likely because the retailer's excess stock buffered the upstream stages. H1 is supported for four of the five error structures.

Table 9: Regression models of the hidden cost change (OLS, HC3 standard errors in parentheses; Models 1 to 3 pool the automation and oversight runs; reference categories: stationary regime, white noise and, in Model 4, full automation)

Predictor	Model 1	Model 2	Model 3	Model 4
Intercept	22.68*** (0.64)	-0.21 (0.63)	34.43*** (0.72)	45.50*** (1.17)
Regime: seasonal	-1.03 (0.89)	-1.03 (0.82)	-1.03* (0.52)	-1.03* (0.47)
Regime: volatile	-11.23*** (0.76)	-11.23*** (0.70)	-11.23*** (0.49)	-11.23*** (0.45)
Error RMS (per 10 units)		19.62*** (0.59)	19.62*** (0.40)	19.62*** (0.36)
Rare spikes			-20.62*** (0.89)	-17.55*** (1.45)
Autocorrelated			-46.49*** (0.72)	-56.12*** (1.19)
Under-forecast bias			-51.32*** (0.75)	-61.65*** (1.24)
Over-forecast bias			-54.79*** (0.76)	-67.50*** (1.25)
Human oversight				-22.14*** (1.27)
Spikes x oversight				-6.14*** (1.56)
Autocorrelated x oversight				19.27*** (1.34)
Under-bias x oversight				20.66*** (1.41)
Over-bias x oversight				25.41*** (1.42)
R squared	0.026	0.177	0.626	0.692
N	9,000	9,000	9,000	9,000

Figure 6: Hidden cost increase by error magnitude and structure (a) and the share of the cost increase borne upstream (b)



Source: simulation experiment of this study (a39/sim39.py); automation condition; panel (b) omits over-forecast bias, whose hidden cost change is negative.

Model 3 adds the error structure with white noise as the reference. All other structures produced significantly less hidden cost than white noise at the same magnitude: rare spikes -20.6, autocorrelated errors -46.5, under-forecast bias -51.3 and over-forecast bias -54.8 per week (all p < 0.001), and the explained variance rose to 0.626. Comparing the two groups of H2 directly, persistent structures raised

hidden cost by 2.6 per week and transient structures by 55.5 (difference controlling for magnitude and regime b = -53.0, SE 0.74, p < 0.001). The difference in visible cost between the groups was negligible (b = -0.18, p = 0.613), so the whole difference in total cost lay upstream. H2 is not supported: the effect has the opposite sign to the one hypothesised. Table 10 repeats the models for the total cost change: white noise and rare spikes remain the most expensive structures, whereas among the persistent structures over-forecast bias, which is costly at the retailer, now comes closest to them.

Table 10: Regression models of the total cost change (OLS, HC3 standard errors in parentheses; same specification as Table 9)

Predictor	Model 1	Model 2	Model 3	Model 4
Intercept	38.98*** (0.80)	-3.24*** (0.69)	30.28*** (0.82)	45.20*** (1.34)
Regime: seasonal	-0.60 (1.11)	-0.60 (0.90)	-0.60 (0.64)	-0.60 (0.55)
Regime: volatile	-17.73*** (0.95)	-17.73*** (0.78)	-17.73*** (0.60)	-17.73*** (0.53)
Error RMS (per 10 units)		36.19*** (0.65)	36.19*** (0.49)	36.19*** (0.42)
Rare spikes			-20.90*** (1.06)	-16.53*** (1.66)
Autocorrelated			-44.36*** (0.86)	-53.94*** (1.38)
Under-forecast bias			-59.85*** (0.95)	-75.96*** (1.55)
Over-forecast bias			-42.45*** (0.83)	-54.38*** (1.36)
Human oversight				-29.85*** (1.42)
Spikes x oversight				-8.74*** (1.78)
Autocorrelated x oversight				19.16*** (1.55)
Under-bias x oversight				32.21*** (1.75)
Over-bias x oversight				23.86*** (1.49)
R squared	0.043	0.368	0.645	0.726
N	9,000	9,000	9,000	9,000

Serial mediation

Figure 7 shows the two mediators. Order distortion was large for white noise and rare spikes and small for the other structures, and upstream bullwhip followed the same pattern, whereas biased forecasts left the manufacturer's order variance close to its benchmark level. Table 11 and Figure 8 report the serial mediation model, estimated on all 9,000 experimental runs with error structure, oversight and regime as covariates. Error magnitude increased order distortion (a1 = 8.87, 95% CI [8.64, 9.10]), order distortion increased upstream bullwhip (d21 = 1.069, 95% CI [1.007, 1.131]), and both mediators increased hidden cost (b1 = 1.923, b2 = 0.731). The serial indirect effect through both mediators was 6.93 (95% CI [6.55, 7.32]), the indirect effect through order distortion alone was 17.06 (95% CI [16.31, 17.84]), and the total indirect effect was 22.14, against a total effect of 19.62. The direct effect was small and negative (-2.52, 95% CI [-2.84, -2.20]), as was the path through bullwhip that bypasses order distortion (-1.85): holding order distortion fixed, larger errors were associated with slightly lower hidden cost, a pattern consistent with buffering by the retailer's safety stock, which grows with the realised errors. The positive and significant serial indirect effect supports H3.

Table 11: Serial mediation of the effect of AI error magnitude on hidden cost (two mediators in sequence, percentile bootstrap with 5,000 resamples)

Path or effect	Estimate	95% bootstrap CI
a1: error RMS -> order distortion	8.87	[8.64, 9.10]
a2: error RMS -> bullwhip ratio	-2.53	[-2.87, -2.18]
d21: order distortion -> bullwhip ratio	1.069	[1.007, 1.131]

Path or effect	Estimate	95% bootstrap CI
b1: order distortion -> hidden cost	1.923	[1.867, 1.977]
b2: bullwhip ratio -> hidden cost	0.731	[0.697, 0.769]
c: total effect of error RMS	19.62	[18.90, 20.37]
c': direct effect of error RMS	-2.52	[-2.84, -2.20]
Indirect: X -> M1 -> Y	17.06	[16.31, 17.84]
Indirect: X -> M1 -> M2 -> Y (serial)	6.93	[6.55, 7.32]
Indirect: X -> M2 -> Y	-1.85	[-2.08, -1.62]
Total indirect effect	22.14	[21.32, 22.98]

Figure 7: Mediators by error structure and magnitude: order distortion (a) and upstream bullwhip ratio (b)

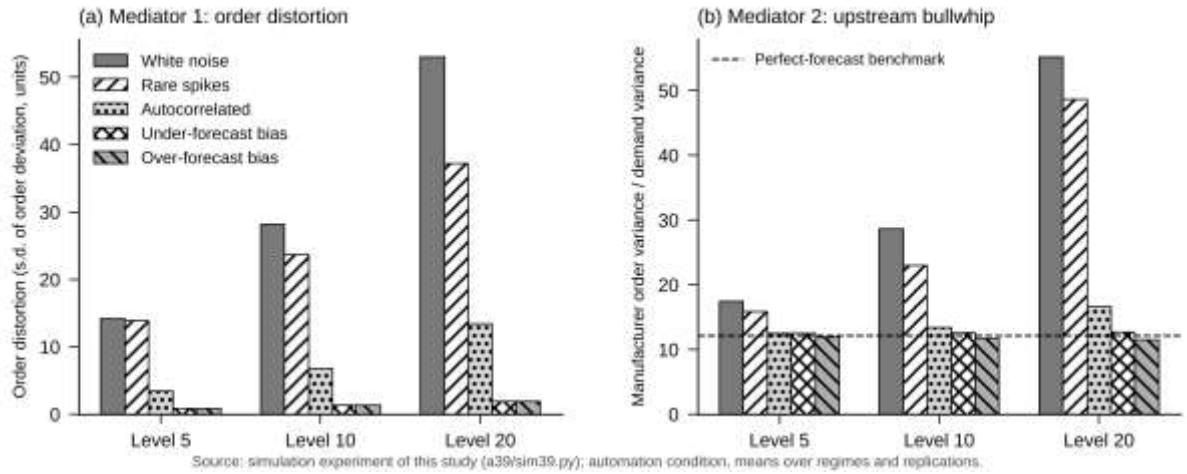


Figure 8: Serial mediation model with path estimates and 95% bootstrap confidence intervals

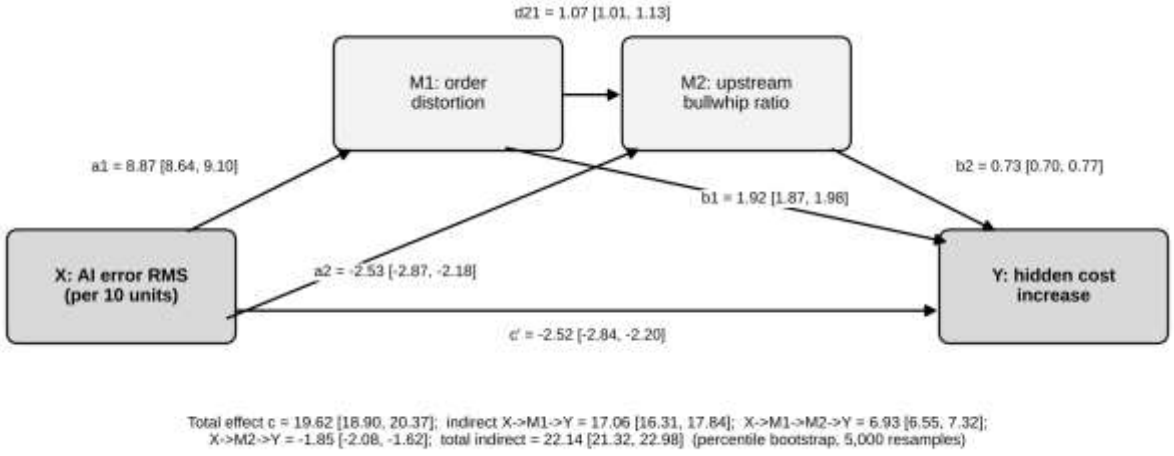


Table 12 estimates the model separately for each structure. The serial indirect effect was largest for rare spikes (14.96) and white noise (11.43), smaller for autocorrelated errors (3.68), and close to zero for the two biases, whose effect on hidden cost was itself close to zero. The mechanism of H3 therefore operates where the hidden cost arises: for errors that change from week to week.

Table 12: Effects of error magnitude on hidden cost by error structure: total, indirect and direct effects with 95% bootstrap confidence intervals (1,000 resamples)

Error structure		Total effect c	X -> M1 -> Y	X -> M1 -> M2 -> Y	X -> M2 -> Y	Direct c'
White noise	Estimate	53.42	42.91	11.43	-1.63	0.71
	95% CI	[52.15, 54.67]	[40.52, 45.30]	[10.16, 12.86]	[-2.67, -0.55]	[-1.15, 2.59]
Rare spikes	Estimate	36.10	30.55	14.96	-5.95	-3.47
	95% CI	[34.56, 37.66]	[27.72, 33.29]	[13.07, 16.95]	[-7.49, -4.47]	[-5.58, -1.35]
Autocorrelated	Estimate	8.65	7.53	3.68	-1.12	-1.44
	95% CI	[8.17, 9.15]	[6.62, 8.40]	[2.96, 4.40]	[-1.70, -0.53]	[-2.22, -0.61]
Under-forecast bias	Estimate	0.43	-1.05	-0.12	0.07	1.52
	95% CI	[0.19, 0.65]	[-1.26, -0.81]	[-0.16, -0.07]	[0.04, 0.11]	[1.35, 1.69]
Over-forecast bias	Estimate	-0.47	1.14	-0.02	-0.04	-1.55
	95% CI	[-0.65, -0.28]	[0.99, 1.33]	[-0.06, 0.01]	[-0.09, 0.00]	[-1.72, -1.39]

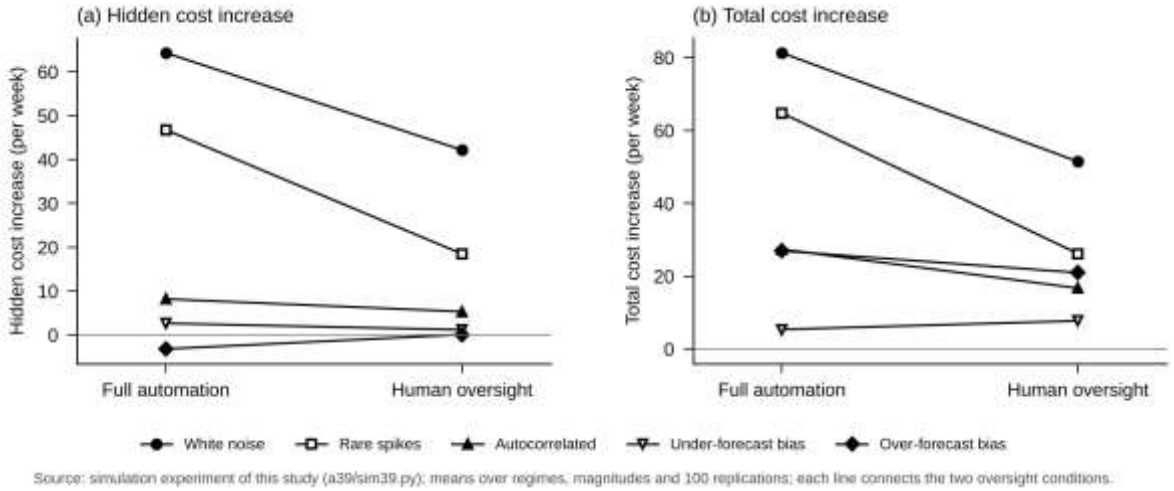
Moderation by human oversight

Table 13 and Figure 9 compare full automation with human oversight. Oversight reduced the hidden cost of rare spikes from 46.8 to 18.5 per week (60%), while overriding the AI forecast in only 4% of the weeks. It reduced the hidden cost of white noise by 34% and of autocorrelated errors by 35%, with override rates of 16% and 17%. For the biases the picture was different. For under-forecast bias oversight lowered hidden cost but raised total cost from 5.4 to 7.7 per week, an increase that arose entirely at the largest magnitude (from 6.4 to 14.6 at 20 units, while total cost fell at 5 and 10 units), most likely because blending with a smoothing forecast that lags demand added its own error. For over-forecast bias oversight lowered total cost by 22% but raised the hidden cost change from -3.2 to 0.1, probably because it removed part of the buffering by the retailer's excess stock. The interactions in Model 4 of Table 9 confirm that the effect of oversight differed by structure; the reduction of hidden cost was strongest for rare spikes, as hypothesised. H4 is supported for four of the five structures, the exception being the hidden cost of over-forecast bias. With a perfect forecast, the oversight rule added 0.85 per week to the benchmark cost, overriding in 1% of the weeks.

Table 13: Hidden and total cost changes under full automation (auto) and human oversight (overs.) by error structure, with the percentage change from automation to oversight (means over regimes and magnitudes, cost units per week)

Error structure	Hidden, auto	Hidden, overs.	Change	p (paired)	Total, auto	Total, overs.	Change	Overrides
White noise	64.3	42.2	-34%	< 0.001	81.3	51.5	-37%	16%
Rare spikes	46.8	18.5	-60%	< 0.001	64.8	26.2	-60%	4%
Autocorrelated	8.2	5.3	-35%	< 0.001	27.4	16.7	-39%	17%
Under-forecast bias	2.7	1.2	-56%	< 0.001	5.4	7.7	+44%	19%
Over-forecast bias	-3.2	0.1	n/a	< 0.001	26.9	20.9	-22%	22%

Figure 9: Moderation by human oversight: hidden (a) and total (b) cost increase under full automation and under the override rule



Robustness

Five variants of the model were run with 40 replications per cell and random numbers independent of the main experiment (Table 14; Figure 10 shows the four variants that affect full automation): a lead time of four weeks, a retail backorder cost of 20, independent demand (phi = 0), a slower upstream smoothing constant (0.1) and a tighter oversight threshold (10%). In every variant white noise and rare spikes produced the largest hidden costs and the biases the smallest, persistent structures produced less hidden cost than transient ones, and oversight reduced the hidden cost of both noise and spikes. The effects were magnified by a longer lead time, which raised the mean hidden cost of transient errors from 55.5 to 181.9 per week, and dampened by slower upstream smoothing (40.0), consistent with the role of the replenishment rules in the mechanism. A tighter oversight threshold increased the reduction for white noise from 34% to 54%. In two variants, however, oversight sharply raised the hidden cost of biased forecasts: with a lead time of four weeks the hidden cost change of under-forecast bias rose from 7.5 to 22.6 and that of over-forecast bias from -6.2 to 24.2, and with independent demand from 2.8 to 16.4 and from -3.1 to 17.5 per week. The value of the override rule therefore depends on the error structure and on the setting, which reinforces the qualification of H4. Table 15 summarises the tests of the four hypotheses.

Table 14: Robustness of the main results to the model parameters (full automation unless stated; cost units per week)

Scenario	Largest hidden costs (descending)	Hidden share	Persistent: hidden	Transient: hidden	Oversight reduction (noise / spikes)
Main experiment	White noise, Rare spikes, Autocorrelated	58%	2.6	55.5	34% / 60%
Retail backorder cost 20	White noise, Rare spikes, Autocorrelated	54%	2.7	56.6	34% / 61%
Independent demand (phi = 0)	White noise, Rare spikes, Autocorrelated	61%	4.4	74.4	26% / 54%
Lead time 4 weeks	Rare spikes, White noise, Autocorrelated	69%	14.4	181.9	29% / 65%
Override threshold 10%	White noise, Rare spikes, Autocorrelated	57%	2.6	55.9	54% / 61%
Upstream smoothing 0.1	White noise, Rare spikes, Autocorrelated	50%	1.7	40.0	44% / 62%

Figure 10: Robustness: mean hidden cost change by error structure in the main experiment and four parameter variants

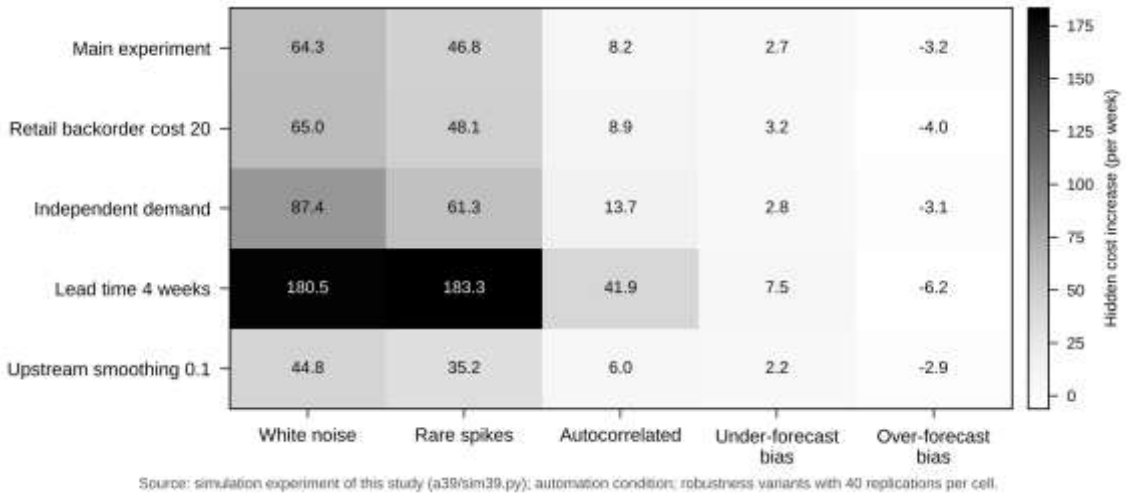


Table 15: Summary of the hypothesis tests

Hypothesis	Test	Result	Decision
H1: AI forecast error raises hidden cost, increasingly with its magnitude	One-sided t test; Model 2 of Table 9	Mean hidden change 23.7, t = 41.5, p < 0.001; b = 19.62 per 10 units	Supported (four of five structures)
H2: at equal accuracy, persistent errors raise hidden cost more than transient errors	Welch t test; regression with level and regime	Persistent 2.6 vs transient 55.5; b = -53.0, p < 0.001	Not supported (reversed)
H3: order distortion and upstream bullwhip serially mediate the effect of error on hidden cost	Serial mediation, 5,000 bootstrap resamples	Serial indirect 6.93 [6.55, 7.32]; total indirect 22.14	Supported
H4: human oversight attenuates hidden cost, most for rare large errors	Paired t tests; interactions in Model 4	Reduction: spikes 60%, noise 34%, autocorrelated 35%	Supported for four of five structures; oversight raised the cost of large biases

Post hoc analysis: text analysis of the research literature

The experiment shows that the cost of AI errors can be large and hidden. To examine how the research literature frames AI in supply chain decisions, we conducted a post hoc text analysis in the spirit of the analysis of social media discussions in earlier work (Grover et al., 2022; Sadeghi R. et al., 2024), applied to research-article titles. The Crossref index was searched with ten strings: artificial intelligence and supply chain decision making, machine learning demand forecasting, forecast error and inventory cost, algorithm aversion, human-AI collaboration in operations, explainable AI in supply chains, AI errors in operations decisions, the bullwhip effect and forecasting, predictive analytics and supply chain risk, and judgmental adjustment of statistical forecasts; the first 100 records of each string were requested in two pages of 50 (900 records were retrieved, because 2 pages failed repeatedly), and after removing duplicates, records other than journal articles and records outside 2000 to 2025, and after removing 17 non-article items such as editorial-board listings, 755 unique titles remained. Titles were tokenised, stop words and generic research terms were removed, frequent compounds (supply chain, artificial

intelligence, machine learning, human-AI) were kept as single terms, and the abbreviation AI was kept although other tokens of fewer than three characters were removed.

The most frequent bigrams (Figure 11) describe applications and methods, led by demand forecasting (in 60 titles), supply chain management (57) and algorithm aversion (50). Among the 373 titles that name AI, machine learning, deep learning, neural networks or algorithms, 111 (30%) used benefit vocabulary (improve, enhance, optimise, efficiency, performance, benefit, value, resilience, accuracy, smart, intelligent) and 98 (26%) used error or risk vocabulary (error, bias, failure, mistake, inaccuracy, risk, uncertainty, aversion, hidden, harm, cost of). Many of the latter come from the search string on algorithm aversion: when titles that qualify only through the word algorithm and the term aversion are set aside, 100 of 300 AI titles (33%) used benefit and 39 (13%) error or risk vocabulary. Only 2 mentioned cost and 1 combined an error term with cost. Across the whole corpus of 755 titles, 5 linked forecast or prediction errors to cost. A latent Dirichlet allocation model (Blei, 2012) was fitted for two to twelve topics. Because the UMass coherence of the ten highest-weight terms varied little across solutions and was highest at the largest number of topics tried (12), the number of topics was set by a parsimony rule, the smallest number whose coherence was within 0.1 of the maximum (Figure 12a); seven topics met the rule by a margin of about 0.001, so the solution should be read as one reasonable partition rather than a unique one. The resulting 7-topic solution (Table 16, Figure 12b) contains topics on AI for sustainable and resilient supply chain management (17.6% of titles), demand forecasting, inventory and cost with machine learning (18.5%), human-AI collaboration and AI errors in decisions (16.6%), the bullwhip effect and decision-making under uncertainty, forecast error and judgmental adjustment, algorithm aversion and risk, and predictive analytics for supply chain risk and resilience. Error terms lead two topics (forecast error and judgmental adjustment; human-AI collaboration), and the cost term leads one (demand forecasting and inventory), but no topic joins error and cost among its twelve leading terms. In this corpus, then, research on AI errors and research on the cost of forecast errors form largely separate conversations, and the cost of AI errors along the chain, the subject of the experiment, is almost absent from the titles. Because the corpus was assembled with search strings chosen for this study, including strings on errors and algorithm aversion, the shares describe this corpus rather than the literature as a whole.

Figure 11: Post hoc text analysis: most frequent bigrams (a) and terms (b) in the harvested research-article titles

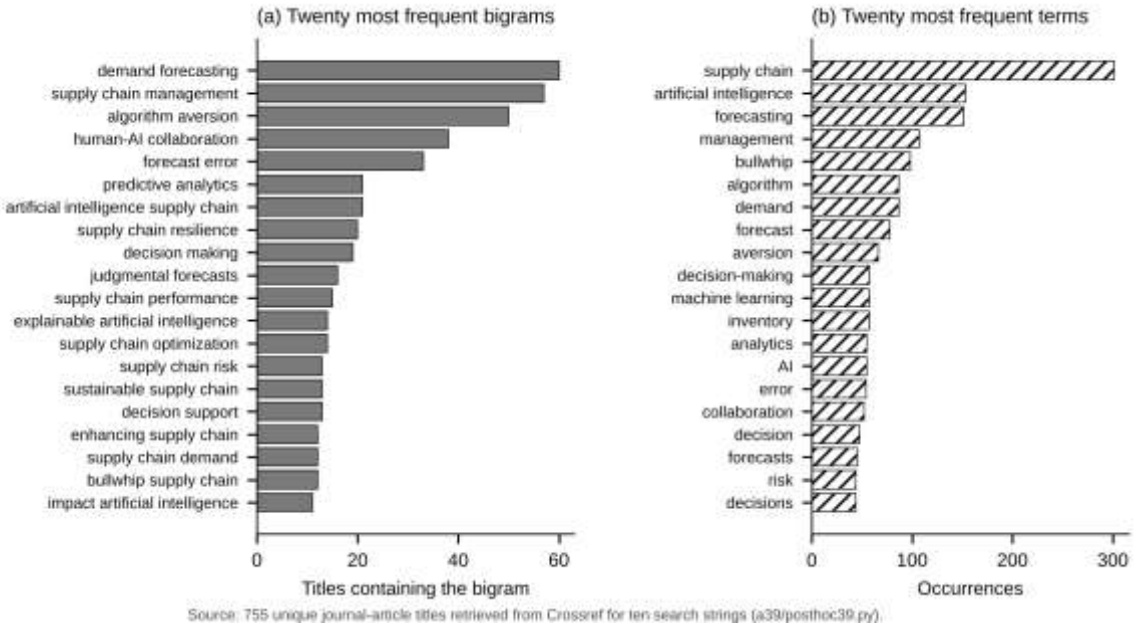
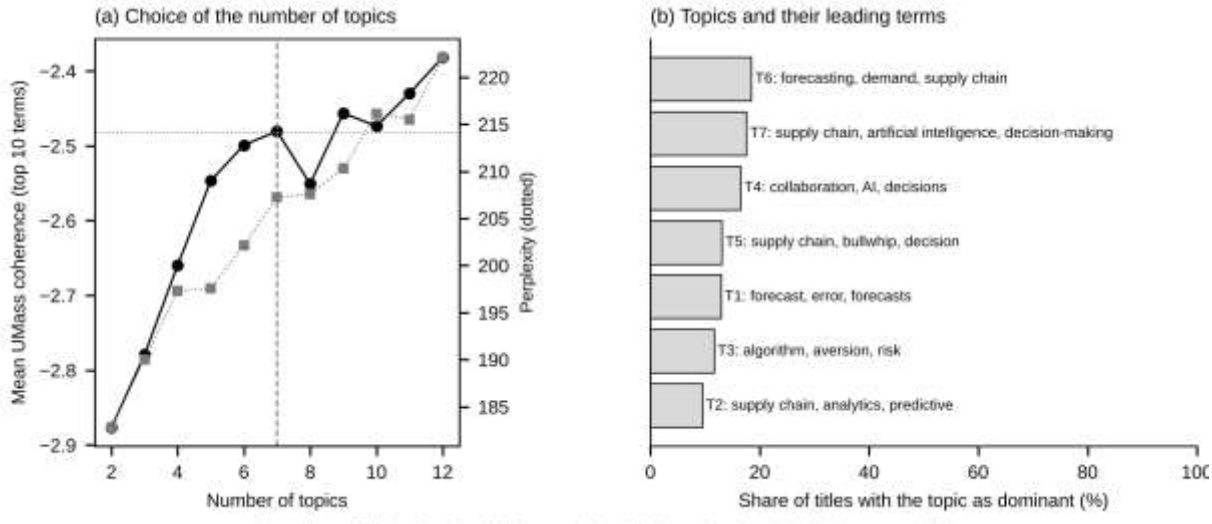


Table 16: Topics of the research-article titles (latent Dirichlet allocation; number of topics chosen by a parsimony rule on UMass coherence)

Topic	Interpretation	Highest-weight terms	Share of titles
1	Forecast error, judgmental adjustment and accuracy	forecast, error, forecasts, judgmental, statistical, adjustment, accuracy, stochastic	13%
2	Predictive analytics and AI for supply chain risk and resilience	supply chain, analytics, predictive, optimization, management, artificial intelligence, enhancing, data	10%
3	Algorithm aversion and risk	algorithm, aversion, risk, forecasting, algorithms, evaluation, appreciation, volatility	12%
4	Human-AI collaboration and AI errors in decisions	collaboration, AI, decisions, artificial intelligence, human-AI, operations, errors, human	17%
5	Bullwhip effect and decision-making under uncertainty	supply chain, bullwhip, decision, impact, making, management, fuzzy, network	13%
6	Demand forecasting, inventory and cost with machine learning	forecasting, demand, supply chain, inventory, bullwhip, machine learning, cost, supply chains	19%
7	AI for sustainable and resilient supply chain management	supply chain, artificial intelligence, decision-making, management, sustainable, resilience, supply chains, performance	18%

Figure 12: Latent Dirichlet allocation: choice of the number of topics (a) and topic shares with their leading terms (b)



Source: latent Dirichlet allocation of the harvested titles (scikit-learn, batch variational inference, seed 39).

DISCUSSION AND CONCLUSIONS

This study asked how large the hidden costs of AI forecast errors in a supply chain are and what determines them. Using a controlled computational experiment, it found that the cost of an AI forecast error is often paid far from the point of decision: for unbiased but unstable errors and for rare large errors, about 70% or more of the cost increase arose at the distributor and the manufacturer, and an evaluation at the retailer would have understated the cost of the error by a factor of three to seven (H1). Equally accurate errors produced very different costs, and contrary to the emphasis on bias in forecast evaluation, persistent errors produced far less hidden cost than transient ones (H2, not supported). The effect of error magnitude on hidden cost ran mainly through order distortion at the decision point and

partly, in series, through bullwhip amplification upstream (H3). A simple human override rule removed much of the hidden cost of rare spikes, but it raised the total cost of large under-forecast biases and, in two robustness variants, the hidden cost of biased forecasts (H4, supported for four of the five structures). The post hoc text analysis showed that research titles on AI in supply chains rarely connect AI errors with their cost.

Theoretical implications

First, the study extends the theory of information distortion (Chen, Drezner, et al., 2000; Lee et al., 1997; Wang & Disney, 2016) to AI forecasts by treating the forecast error as a structured signal that the replenishment rules transmit and amplify. The reversal of H2 follows from this view. An order-up-to policy passes on the change in the forecast, not its level, multiplied by the protection interval. A constant bias changes the forecast once and then shifts the order-up-to level permanently, so it can raise the inventory of the stage that uses it (over-forecast bias), or its backorders where the safety stock does not absorb it, but leaves the variability of its orders, and hence the upstream stages, almost untouched. An error that changes every week, by contrast, is passed on in full every week and is then amplified by the smoothing and safety-stock rules upstream (Chen, Ryan, et al., 2000; Dejonckheere et al., 2003). The relevant property of an AI error for the chain is therefore its variability over time, which accuracy measures that average squared or absolute errors do not capture. This qualifies the attention that forecast evaluation gives to bias (Fildes & Kingsman, 2011; Sanders & Graman, 2009): over-forecast bias produced the largest cost increase at the stage that used the forecast, whereas instability produced the largest cost increase for the chain.

Second, the study contributes to the literature on evaluating forecasts by their decision consequences (Gneiting, 2011; Kourentzes et al., 2020; Syntetos et al., 2010) and to work on decision-focused learning (Bertsimas & Kallus, 2020; Elmachtoub & Grigas, 2022). That literature has shown that accuracy and inventory performance can diverge at the stage that uses the forecast. The present results confirm the divergence at that stage and show that it also holds for the chain as a whole: RMSE and MAE together explained 18% of the variation in the retailer's own cost increase and 31% of the variation in the total cost increase, and the MAE ranked the second most expensive error structure, rare spikes, as the most accurate. Moreover, the part of the cost that accuracy misses is located upstream, where the evaluation of the forecast does not look. A complete evaluation of an AI forecast must therefore include a model of how its errors propagate, not only a loss function at the decision point.

Third, the study adds a cost perspective to the research on human interaction with algorithms. That research has concentrated on whether people use or override algorithms and with what effect on the decision at hand (Caro & Sáez de Tejada Cuenca, 2023; Dietvorst et al., 2015; Kesavan & Kushwaha, 2020; Logg et al., 2019; Sun et al., 2022). The moderation results show that the value of human oversight depends on the structure of the errors it is meant to catch: an override rule that is triggered by large deviations is well suited to gross errors, removes only about a third of the hidden cost of unstable and slowly drifting errors, and can raise cost when the forecast is biased, probably because the benchmark it falls back on lags demand. Oversight should therefore be studied as a design choice matched to the error profile of a model, consistent with the view that human and AI judgement are complements whose value depends on the task (Brau et al., 2023; Khosrowabadi et al., 2022; Raisch & Krakowski, 2021).

Managerial implications

For managers who deploy AI forecasts, the results suggest four practices. First, evaluate an AI forecast by the cost it causes across the supply chain, or at least by the variability of the orders it induces, not only by its RMSE or MAE at the point of use, even when that accuracy is measured properly out of sample (Tashman, 2000). In the experiment, the variance of the retailer's orders relative to the variance of demand, which a firm can observe, correlated with hidden cost at $r = 0.84$, against 0.24 for the RMSE (order distortion relative to a perfect forecast, which cannot be observed in practice, correlated at 0.95), so a measure of this kind would have flagged the expensive error structures that accuracy measures missed. Second, when choosing between models, prefer the one whose errors are stable over time to one that is slightly more accurate but unstable, and treat occasional large errors as a priority even when they hardly move the average error. Third, design oversight for the error profile of the model: a deviation-triggered review is cheap and effective against gross errors, but persistent biases require a

different control, such as bias monitoring with tracking signals, and in our model blending with a lagging benchmark made large biases more expensive. Fourth, because the hidden cost falls on suppliers and on internal functions such as production, the evaluation of an AI forecast is a matter for the whole chain; sharing forecast and order-variability metrics with partners, and including their costs in the business case for an AI system, would make the hidden cost visible.

Limitations and future research

The findings rest on a stylised model, and their limits should be stated plainly. The AI forecast was represented by a perfect-information forecast plus an injected error; this isolates the error structure but does not reproduce the joint structure of the errors of any specific learning algorithm, and future work should repeat the analysis with trained models on real demand data, such as the retail data of forecasting competitions (Makridakis et al., 2022). The chain is serial with three stages, fixed lead times, linear costs and order-up-to policies with exponential-smoothing forecasts upstream; networks with several customers and suppliers, capacity limits, ordering costs and upstream stages that use AI themselves may behave differently, although the robustness checks show that the grouping of error structures (transient errors above persistent ones) held across lead times, cost ratios, demand autocorrelation and smoothing constants. The oversight condition is a deterministic rule rather than the behaviour of real planners, who adjust forecasts with their own biases (Fildes et al., 2009; Kremer et al., 2011); a behavioural experiment in which participants review AI forecasts with different error structures would test whether people detect and correct the expensive structures. The cost parameters are illustrative, so the absolute cost figures should not be read as estimates for any industry, whereas the relative findings (the ordering of structures, the share borne upstream and the mediation) are the substance of the results. Finally, the post hoc text analysis uses titles, not full texts, and up to the first 100 records of ten searches in one index, so it describes the framing of the literature only coarsely. Future research could measure the hidden cost of AI errors in firm data where upstream costs are observable, develop evaluation metrics that combine accuracy with order stability, and study how incentives and contracts can allocate the hidden cost to the party that deploys the model.

Conclusion

AI forecasts are judged where they are used, but their errors are paid for along the chain. In a controlled experiment, errors of identical accuracy produced cost increases that differed several-fold, the most expensive errors were unbiased but unstable, and most of their cost fell on upstream stages that an evaluation at the decision point does not see. The mechanism is the distortion of orders and its amplification upstream, and the remedy is not only better accuracy but forecasts whose errors are stable, oversight designed for the errors a model actually makes, and evaluation by the cost a forecast causes across the supply chain.

REFERENCES

[1]. Ali, M. M., Boylan, J. E., & Syntetos, A. A. (2012). Forecast errors and inventory performance under forecast information sharing. *International Journal of Forecasting*, *28*(4), 830-841. https://doi.org/10.1016/j.ijforecast.2010.08.003

[2]. Aminul Islam, M., & Md, A. (2023). Adaptive Class-Imbalance Learning for High-Sensitivity Fraud Detection in Large-Scale Digital Transaction Networks. *International Journal of Business and Economics Insights*, *13*(09), 01-33. https://doi.org/10.63125/bmxwv741

[3]. Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2019). Integrating human judgement into quantitative forecasting methods: A review. *Omega*, *86*, 237-252. https://doi.org/10.1016/j.omega.2018.07.012

[4]. Asheke, M., & Tashriq Ahmed Bin, A. (2022). Federated Learning and Differential Privacy for Cross-Firm Construction-Materials Supply-Chain Disruption Forecasting: A Quantitative Analysis of Privacy-Utility Trade-Offs. *American Journal of Advanced Technology and Engineering Solutions*, *12*(09), 01-40. https://doi.org/10.63125/g4at4237

[5]. Ban, G.-Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, *67*(1), 90-108. https://doi.org/10.1287/opre.2018.1757

[6]. Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

[7]. Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: State of the art and future research directions. *International Journal of Production Research*, *57*(7), 2179-2202. https://doi.org/10.1080/00207543.2018.1530476

[8]. Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, *66*(3), 1025-1044. https://doi.org/10.1287/mnsc.2018.3253

[9]. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84. https://doi.org/10.1145/2133806.2133826

[10]. Brau, R., Aloysius, J., & Siemsen, E. (2023). Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics. *Journal of Operations Management*, *69*(6), 965-982. https://doi.org/10.1002/joom.1257

[11]. Bray, R. L., & Mendelson, H. (2012). Information transmission and the bullwhip effect: An empirical investigation. *Management Science*, *58*(5), 860-875. https://doi.org/10.1287/mnsc.1110.1467

[12]. Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220-239. https://doi.org/10.1002/bdm.2155

[13]. Cachon, G. P., Randall, T., & Schmidt, G. M. (2007). In search of the bullwhip effect. *Manufacturing & Service Operations Management*, *9*(4), 457-479. https://doi.org/10.1287/msom.1060.0149

[14]. Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140-1154. https://doi.org/10.1016/j.ejor.2006.12.004

[15]. Caro, F., & Sáez de Tejada Cuenca, A. (2023). Believing in analytics: Managers' adherence to price recommendations from a DSS. *Manufacturing & Service Operations Management*, *25*(2), 524-542. https://doi.org/10.1287/msom.2022.1166

[16]. Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, *46*(3), 436-443. https://doi.org/10.1287/mnsc.46.3.436.12069

[17]. Chen, F., Ryan, J. K., & Simchi-Levi, D. (2000). The impact of exponential smoothing forecasts on the bullwhip effect. *Naval Research Logistics*, *47*(4), 269-286. https://doi.org/10.1002/(sici)1520-6750(200006)47:4<269::aid-nav1>3.0.co;2-q

[18]. Choi, T. M., Wallace, S. W., & Wang, Y. (2018). Big data analytics in operations management. *Production and Operations Management*, *27*(10), 1868-1883. https://doi.org/10.1111/poms.12838

[19]. Croson, R., & Donohue, K. (2006). Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Science*, *52*(3), 323-336. https://doi.org/10.1287/mnsc.1050.0436

[20]. Dejonckheere, J., Disney, S. M., Lambrecht, M. R., & Towill, D. R. (2003). Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research*, *147*(3), 567-590. https://doi.org/10.1016/s0377-2217(02)00369-7

[21]. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114-126. https://doi.org/10.1037/xge0000033

[22]. Disney, S. M., & Towill, D. R. (2003). On the bullwhip and inventory variance produced by an ordering policy. *Omega*, *31*(3), 157-167. https://doi.org/10.1016/s0305-0483(03)00028-8

[23]. Dolgui, A., & Ivanov, D. (2020). Exploring supply chain structural dynamics: New disruptive technologies and disruption risks. *International Journal of Production Economics*, *229*, 107886. https://doi.org/10.1016/j.ijpe.2020.107886

[24]. Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, *68*(1), 9-26. https://doi.org/10.1287/mnsc.2020.3922

[25]. 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. https://doi.org/10.1016/j.ijforecast.2008.11.010

[26]. Fildes, R., & Kingsman, B. (2011). Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society*, *62*(3), 483-500. https://doi.org/10.1057/jors.2010.40

[27]. Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283-1318. https://doi.org/10.1016/j.ijforecast.2019.06.004

[28]. Gazi Yeasin, A. (2024a). Designing Trustworthy AI Output Interfaces: A UX Framework for Graceful Handling of Generative Model Uncertainty in Consumer Applications. *American Journal of Scholarly Research and Innovation*, *14*(07), 01–40. https://doi.org/10.63125/4m5fjb12

[29]. Gazi Yeasin, A. (2024b). A Quantitative Evaluation of Behavioral Design Patterns on User Engagement and Activation in Large-Scale Campus Software Products. *American Journal of Interdisciplinary Studies*, *14*(06), 01-40. https://doi.org/10.63125/x3d5k517

[30]. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

[31]. Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746-762. https://doi.org/10.1198/jasa.2011.r10138

[32]. Grover, P., Kar, A. K., & Dwivedi, Y. K. (2022). Understanding artificial intelligence adoption in operations management: Insights from the review of academic literature and social media discussions. *Annals of Operations Research*, *308*(1-2), 177-213. https://doi.org/10.1007/s10479-020-03683-9

[33]. Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the New millennium. *Communication Monographs*, *76*(4), 408-420. https://doi.org/10.1080/03637750903310360

[34]. Helo, P., & Hao, Y. (2022). Artificial intelligence in operations management and supply chain management: An exploratory case study. *Production Planning & Control*, *33*(16), 1573-1590. https://doi.org/10.1080/09537287.2021.1882690

[35]. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[36]. Ivanov, D., Dolgui, A., & Sokolov, B. (2019). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, *57*(3), 829-846. https://doi.org/10.1080/00207543.2018.1488086

[37]. Kazi Mohammad Khalid, A., & Sumaiya Sadmeen, M. (2022). Geospatial Artificial Intelligence for Urban Flood Susceptibility and Drainage Resilience: A Prisma 2020 Systematic Review and Bibliometric Meta-Synthesis. *Review of Applied Science and Technology*, *12*(02), 01–51. https://doi.org/10.63125/epqdvy23

[38]. Kesavan, S., & Kushwaha, T. (2020). Field experiment on the profit implications of merchants' discretionary power to override data-driven decision-making tools. *Management Science*, *66*(11), 5182-5190. https://doi.org/10.1287/mnsc.2020.3743

[39]. Khosrowabadi, N., Hoberg, K., & Imdahl, C. (2022). Evaluating human behaviour in response to AI recommendations for judgemental forecasting. *European Journal of Operational Research*, *303*(3), 1151-1167. https://doi.org/10.1016/j.ejor.2022.03.017

[40]. Kleijnen, J. P. C. (2005). An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research*, *164*(2), 287-300. https://doi.org/10.1016/j.ejor.2004.02.005

[41]. Kourentzes, N., Trapero, J. R., & Barrow, D. K. (2020). Optimising forecasting models for inventory planning. *International Journal of Production Economics*, *225*, 107597. https://doi.org/10.1016/j.ijpe.2019.107597

[42]. Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, *57*(10), 1827-1843. https://doi.org/10.1287/mnsc.1110.1382

[43]. Lawrence, 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. https://doi.org/10.1016/j.ijforecast.2006.03.007

[44]. Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science*, *43*(4), 546-558. https://doi.org/10.1287/mnsc.43.4.546

[45]. Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

[46]. Maishar Mostari, Z. (2024). Uncertainty-Calibrated Explainable AI for Cybersecurity Alert Prioritization: Improving Detection Reliability and Human Decision-Making. *American Journal of Scholarly Research and Innovation*, *14*(12), 01–40. https://doi.org/10.63125/an0pcq39

[47]. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, *38*(4), 1346-1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

[48]. Mashfiqur, R. (2024). Anomaly Detection and Process Mining for Identifying Financial Irregularities and Strengthening Procurement Controls. *American Journal of Advanced Technology and Engineering Solutions*, *14*(09), 91-140. https://doi.org/10.63125/hmg5ta84

[49]. Mashfiqur, R., & Md Jamil, A. (2023). Machine Learning-Based Cost Variance Prediction for Early Detection of Budget Overruns in Procurement-Intensive Operations. *Review of Applied Science and Technology*, *13*(12), 142–199. https://doi.org/10.63125/7g92wd03

[50]. Masrul, H., & Mashfiqur, R. (2024). Unified Analytical Modeling for Cryptocurrency Price Volatility Forecasting and Market-Manipulation Detection in U.S. Digital Asset Markets. *American Journal of Advanced Technology and Engineering Solutions*, *14*(11), 01-40. https://doi.org/10.63125/nqgh7934

[51]. Masrul, H., & Mukut Kanti, B. (2024). AI-Driven Project Portfolio and Business Analytics for Forecasting Cost Overruns, Schedule Delays, Resource Bottlenecks, and Enterprise Performance Across U.S. Industries. *Review of Applied Science and Technology*, *14*(12), 14–54. https://doi.org/10.63125/x7te4644

[52]. Md, F., & Kynatun, M. (2023). A Quantitative Assessment of Cybersecurity Risk Scoring Models for Enterprise Database Systems in Cloud-Native Environments. *Review of Applied Science and Technology*, *13*(12), 50–103. https://doi.org/10.63125/gk2krm47

[53]. Md Hasinur, R. (2024). Human in the Loop Artificial Intelligence for Loan Document Verification and Credit Risk Reporting: A Bibliometric and Science Mapping Analysis. *American Journal of Interdisciplinary Studies*, *14*(06), 41-71. https://doi.org/10.63125/8x359j95

[54]. Md Maksudul, H. (2024). Supplier-Risk Scoring and Procurement Resilience for Distributed Athletic-Apparel and Equipment Supply Networks. *American Journal of Advanced Technology and Engineering Solutions*, *14*(09), 44-90. https://doi.org/10.63125/a9fmgj50

[55]. Md Maksudul, H., & Md Takbir Hossen, S. (2023). Predictive Inventory Optimization for Reducing Stockouts and Excess Inventory in U.S. Sports and Apparel Operations. *Review of Applied Science and Technology*, *13*(12), 104–141. https://doi.org/10.63125/8djn1j48

[56]. Md Maksudul, H., & Md Tazul Islam, C. (2022). Machine Learning-Based Cash-Flow Forecasting for Improving Financial Planning in U.S. Small and Mid-Sized Businesses. *International Journal of Scientific Interdisciplinary Research*, *12*(12), 01–38. https://doi.org/10.63125/w3pgjx97

[57]. Md Mohsin, B. (2024). Artificial Intelligence for Medication Safety and Nursing Documentation in Emergency Departments: A Bibliometric Analysis. *American Journal of Interdisciplinary Studies*, *14*(10), 01-34. https://doi.org/10.63125/2krdb618

[58]. Md Mohsin, B., & Jannatul, F. (2021). Emergency Nursing Triage and Timely Escalation of Care: A Quantitative Clinical Audit. *International Journal of Scientific Interdisciplinary Research*, *11*(07), 01–33. https://ijsir.org/index.php/IJSIR/article/view/138

[59]. Md Mohsin, B., & Mst Kaniz, F. (2023). AI-Enabled Clinical Decision Support for Standardized Nurse-Led Emergency Care Pathways: A Multisite Mixed-Methods Implementation Study. *Review of Applied Science and Technology*, *13*(08), 01–33. https://doi.org/10.63125/ac09sa31

[60]. Md Tazul Islam, C. (2024). Explainable AI For Financial and Operational Decision Support: Improving Transparency and Managerial Trust. *American Journal of Advanced Technology and Engineering Solutions*, *14*(12), 01-38. https://doi.org/10.63125/kwfvzw68

[61]. Md. Mainuddin, F. (2024). Quantitative Structural Retrofit Assessment Models for Strengthening Existing Steel Buildings Under Increased Load Demands. *Review of Applied Science and Technology*, *3*(04), 325–366. https://doi.org/10.63125/yyqnte84

[62]. Md. Mamunur, R. (2024). Predictive Analytics for Cloud Service Reliability A Comparative Evaluation of System Performance and Disaster-Recovery Metrics. *Review of Applied Science and Technology*, *14*(12), 55–91. https://rast-journal.org/index.php/RAST/article/view/171

[63]. Md. Mamunur, R., & Md, A. (2023). Cloud Migration Readiness and Operational Risk in U.S. Organizations: A Quantitative Analysis of Infrastructure and Business Continuity Indicators. *American Journal of Scholarly Research and Innovation*, *13*(07), 01-27. https://doi.org/10.63125/28jkj206

[64]. 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. https://doi.org/10.1016/j.ejor.2018.10.028

[65]. Pournader, M., Ghaderi, H., Hassanzadegan, A., & Fahimnia, B. (2021). Artificial intelligence applications in supply chain management. *International Journal of Production Economics*, *241*, 108250. https://doi.org/10.1016/j.ijpe.2021.108250

[66]. Prak, D., Teunter, R., & Syntetos, A. (2017). On the calculation of safety stocks when demand is forecasted. *European Journal of Operational Research*, *256*(2), 454-461. https://doi.org/10.1016/j.ejor.2016.06.035

[67]. Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879-891. https://doi.org/10.3758/brm.40.3.879

[68]. Punnaiah Chowdary, R. (2021). GC-MS Strategies for Identification of Trace Impurities in Active Pharmaceutical Ingredients. *American Journal of Advanced Technology and Engineering Solutions*, *11*(05), 01-35. https://ajates-scholarly.com/index.php/ajates/article/view/131

[69]. Punnaiah Chowdary, R., & Md, A. (2020). Practical HPLC Method Validation Under ICH Q2(R2) and Q14: An Analytical Quality Framework. *International Journal of Business and Economics Insights*, *10*(03), 01–36. https://ijbei-journal.org/index.php/ijbei/article/view/73

[70]. Punnaiah Chowdary, R., & Md Jamil, A. (2020). Risk-Based Lifecycle Management of Chromatographic Methods for Generic API Quality: A Systematic Review. *Review of Applied Science and Technology*, *10*(09), 01–41. https://doi.org/10.63125/r0xnar90

[71]. Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141. https://doi.org/10.1007/s11747-019-00710-5

[72]. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, *46*(1), 192-210. https://doi.org/10.5465/amr.2018.0072

[73]. Riahi, Y., Saikouk, T., Gunasekaran, A., & Badraoui, I. (2021). Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions. *Expert Systems with Applications*, *173*, 114702. https://doi.org/10.1016/j.eswa.2021.114702

[74]. Sadeghi R., K., Ojha, D., Kaur, P., Mahto, R. V., & Dhir, A. (2024). Explainable artificial intelligence and agile decision-making in supply chain cyber resilience. *Decision Support Systems*, *180*, 114194. https://doi.org/10.1016/j.dss.2024.114194

[75]. Sadia, A., & Md. Mosheur, R. (2023). Consumer Trust, Perceived Authenticity, and Personalized Digital Marketing: A Quantitative Decision-Science Study. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 252–295. https://doi.org/10.63125/cajkqt60

[76]. Sanders, N., & Graman, G. (2009). Quantifying costs of forecast errors: A case study of the warehouse environment. *Omega*, *37*(1), 116-125. https://doi.org/10.1016/j.omega.2006.10.004

[77]. Seaman, B. (2018). Considerations of a retail forecasting practitioner. *International Journal of Forecasting*, *34*(4), 822-829. https://doi.org/10.1016/j.ijforecast.2018.03.001

[78]. Sharmin, A., & Mst Kaniz, F. (2023). "My Pain Was Not Taken Seriously" Medical Gaslighting As an Expectancy Violation in Women's Healthcare. *American Journal of Health and Medical Sciences*, *13*(06), 55–85. https://ajhmsjournal.com/index.php/ajhms/article/view/20

[79]. Sharmin, A., & Nishat, J. (2023). Communication Accommodation and Patient Trust in Healthcare and Chronic Illness Care. *Review of Applied Science and Technology*, *13*(08), 34–72. https://doi.org/10.63125/6p0xwb79

[80]. Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, *35*(3), 553-572. https://doi.org/10.2307/23042796

[81]. Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, *61*(4), 66-83. https://doi.org/10.1177/0008125619862257

[82]. Songeta, D. (2024). Human-In-The-Loop Artificial Intelligence for Account Reconciliation and Financial-Reporting Transparency: A Bibliometric and Science-Mapping Analysis. *American Journal of Scholarly Research and Innovation*, *14*(12), 41–72. https://doi.org/10.63125/jamadx91

[83]. Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, *35*(3), 321-339. https://doi.org/10.1287/mnsc.35.3.321

[84]. Sumaiya, M. (2024). Hospital Wastewater as A Reservoir of Antimicrobial Resistance Genes: Implications for Public Health and Genomic Surveillance in The United States. *Review of Applied Science and Technology*, *3*(04), 448–508. https://doi.org/10.63125/ahvas976

[85]. Sumaiya Sadmeen, M., & Kazi Mohammad Khalid, A. (2023). Multi-Source Remote Sensing and Dem-Derived Hydrogeomorphic Indicators for High-Resolution Urban Flood Susceptibility Mapping: A Spatial Cross-Validation Framework. *Journal of Sustainable Development and Policy*, *13*(07), 01-46. https://doi.org/10.63125/z4xt7k38

[86]. Sun, J., Zhang, D. J., Hu, H., & Van Mieghem, J. A. (2022). Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science*, *68*(2), 846-865. https://doi.org/10.1287/mnsc.2021.3990

[87]. Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*(1), 1-26. https://doi.org/10.1016/j.ejor.2015.11.010

[88]. 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. https://doi.org/10.1016/j.ijforecast.2009.05.016

[89]. Tahmedur, R. (2024). Machine-Learning-Assisted Project Cost Estimation for Technology-Intensive Engineering Programs. *International Journal of Scientific Interdisciplinary Research*, *14*(12), 01–54. https://doi.org/10.63125/7gfgm030

[90]. Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, *16*(4), 437-450. https://doi.org/10.1016/s0169-2070(00)00065-0

[91]. Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P., & Fischl, M. (2021). Artificial intelligence in supply chain management: A systematic literature review. *Journal of Business Research*, *122*, 502-517. https://doi.org/10.1016/j.jbusres.2020.09.009

[92]. Trapero, J. R., Cardós, M., & Kourentzes, N. (2019). Empirical safety stock estimation based on kernel and GARCH models. *Omega*, *84*, 199-211. https://doi.org/10.1016/j.omega.2018.05.004

[93]. Wang, X., & Disney, S. M. (2016). The bullwhip effect: Progress, trends and directions. *European Journal of Operational Research*, *250*(3), 691-701. https://doi.org/10.1016/j.ejor.2015.07.022

[94]. Zhang, X. (2004). Evolution of ARMA demand in supply chains. *Manufacturing & Service Operations Management*, *6*(2), 195-198. https://doi.org/10.1287/msom.1040.0042