

# ***Research on SKU-Level Inventory Health Monitoring and Dynamic Replenishment Strategies***

**Qian Lin**

*W. P. Carey School of Business, Arizona State University, Mesa, 85287, Arizona, U.S*

**Keywords:** Inventory health; Inventory control; Dynamic replenishment; Probabilistic demand forecasting; Demand truncation

**Abstract:** Addressing the issues of retail stockouts leading to truncated sales and the inability of fixed-probability replenishment to promptly reflect SKU inventory risk, this study examines the incremental decision-making value of inventory status information based on probabilistic demand forecasting. Using FreshRetailNet-50K data, the study employs TimesNet to recover potential demand and utilizes LightGBM quantile regression to predict demand during the protection period, classifying states into shortage, healthy, and excess. Under the same forecasting environment, it compares health-triggered and probability-driven replenishment. Results show that the three states maintain a stable risk ranking; compared to B2, the proposed fill rate increases by 0.37 percentage points, average inventory increases by 1.60%, and the inventory difference after service matching is  $-0.06$  NDU, with confidence intervals crossing zero. The incremental effect mainly occurs in SKUs with high volatility and long lead times. This study demonstrates that inventory health information has conditional replenishment value and can serve as an interpretable risk calibration interface connecting probabilistic forecasting and replenishment rules.

## **1 Introduction**

### **1.1 Research Background and Problem Statement**

Retail inventory management requires striking a balance between product availability and inventory holding. In fresh produce and fast-moving consumer goods (FMCG) scenarios, promotions, holidays, and short-term demand fluctuations continuously alter the actual demand for SKUs. While fixed safety stock is easy to implement, it can lead to stockouts or overstocking after demand changes. Therefore, replenishment strategies must not only focus on future demand but also assess the current risk level of inventory and adjust replenishment accordingly.

Actual sales are also limited by inventory availability. When goods are out of stock, unmet demand cannot be fully recorded in sales records. Directly using such truncated sales may underestimate future demand and further affect inventory decisions [1]. Probabilistic forecasting can describe demand uncertainty, but higher forecast accuracy does not automatically translate into

better inventory performance. The actual effect still depends on factors such as inventory policy, demand structure, and lead time [2].

Based on this question, this study limits its scope to "probabilistic demand—inventory health—dynamic replenishment," without extending to allocation, pricing, multi-level inventory, or supplier decisions. This study focuses on whether inventory status can still provide additional decision-making value after probabilistic demand information has already been incorporated into replenishment decisions.

## 1.2 Related Research and Research Questions

Existing research has gradually shifted from simply comparing forecast errors to examining the actual role of forecast information in inventory decisions. For intermittent demand, existing research has used hybrid LSTM to improve forecast performance [3]; probabilistic forecasting research has combined distribution information to more completely describe demand uncertainty [4]. These works have provided a foundation for characterizing demand risk, but how forecast information further affects replenishment decisions still needs to be analyzed separately.

Another type of research directly incorporates demand information into ordering decisions. Supervised learning has been used for joint modeling of forecasting and inventory[5], and price and demand uncertainties have also been incorporated into replenishment decisions[6]. In dynamic inventory environments, service level dependence on demand changes control strategies[7], DRL can handle inventory optimization under non-stationary demand[8], and Deep Controlled Learning further focuses on learning-based inventory control[9]. Demand forecasting can also be directly incorporated into perishable goods ordering policies[10]. However, these methods often use forecasting, state, or policy learning information simultaneously, so the effects of different information are still not easily separated and identified.

This study does not construct a comprehensive health score, but instead defines a two-sided risk state based on the relationship between inventory location and the demand distribution during the protection cycle[11-12]. The focus is not on adding a new forecasting algorithm, but on examining whether inventory location can still provide additional decision-making information once the future demand distribution is known. B2 and Proposed use the same forecasting model and inventory environment; the only difference is whether shortage risk is incorporated into replenishment feedback[13-15]. From a control structure perspective, both Proposed and the classic (s,S) strategy have threshold triggering characteristics, but the meanings of the thresholds are different. In this study, 0.80 and 0.95 are not fixed inventory levels  $s$  and  $S$ , but rather risk quantile levels that change with the demand distribution during the protection cycle on the decision date; the health state still executes B2 at the 0.90 quantile. This study does not interpret this threshold switching as a new ordering mechanism, but rather transforms the state of inventory location relative to the future demand distribution into an interpretable feedback to identify its incremental value on a stronger probability baseline[16].

To address the aforementioned issues, this study sets forth two research questions. RQ1 focuses on whether inventory health status can form a stable forward risk stratification on a reference trajectory without using health feedback; RQ2 focuses on whether health-triggered replenishment can improve the service-inventory tradeoff of fixed-probability replenishment under the same forecast and inventory conditions. Demand fluctuations, lead times, demand recovery, and threshold changes are only used to test the applicability and robustness of the method[17].

## 2 Research Methods

### 2.1 Data and Protection Cycle Demand Forecast

This study uses FreshRetailNet-50K v1.0 data and takes "store-SKU-day" as the basic unit of analysis. The data contains 50,000 90-day series, covering 898 stores and 18 cities, and provides information on sales, stockouts, and promotions, providing a unified data foundation for demand recovery, probability forecasting, and inventory status assessment.

Before modeling, this study examined data duplication and temporal continuity. Lagged and rolling features only used information available before the decision date, and promotional and calendar variables that could be known in advance were included in the model; the initial 14 days were only used to form a historical window and did not participate in subsequent performance evaluation.

The data was divided chronologically. Days 1-55 were used for development, days 56-69 for validation, and days 70-90 were divided into three 7-day extended windows. Each window only incorporated the training data already implemented at that time, without re-searching for parameters. Finally, this study used 90 days of information to fit the model and tested it on the official 7-day eval dataset.

When there is a shortage, observed sales  $Y_{i,t}$  may be lower than potential demand  $D_{i,t}$ , that is :

$$Y_{i,t} \leq D_{i,t}.$$

To address potential demand during stockout periods, this study employs the official FreshRetailNet TimesNet workflow for recovery. This workflow only replaces sales figures from the stockout period, retaining the actual sales figures when stock is back in stock, and restricts the recovered values to non-negative. To minimize the impact of implementation differences, the experiment uses a fixed snapshot of the official code and runtime dependencies.

Demand recovery is only used to correct for sales cutoff and is not considered additional strategy information. To determine whether this treatment affects the main conclusions, the robustness analysis includes an additional condition of "no demand recovery".

The replenishment system is checked daily, with the check cycle recorded as  $R = 1$  and the replenishment lead time as  $L$ . Therefore, the protection period is :

$$P = R + L.$$

This study directly constructs  $P$  the cumulative demand for future days :

$$D_{i,t}^P = \sum_{k=1}^P D_{i,t+k},$$

This study directly predicts the cumulative demand for future replenishment protection cycles, rather than simply adding the marginal quantiles of different dates. This approach ensures that the forecast target aligns with the actual replenishment protection cycle and reduces explanatory bias caused by the accumulation of quantiles across days.

LightGBM 4.3.0 quantile regression estimated the 0.75, 0.80, 0.85, 0.90, and 0.95 quantiles, respectively. Model inputs included demand lags, rolling statistics, cross-correlation (CV), historical stockouts, and predictable information. Parameters were set to learning rate=0.05, num\_leaves=31, n\_estimators≤300, early stopping=30, and seed=2026; if quantile crossovers occurred, they were rearranged in ascending order. The main experiment was set to  $L=1$ ; changing  $L$  reconstructed the cumulative demand labels.

## 2.2 Inventory Health Monitoring and Dynamic Replenishment Strategy

$OH_{i,t}$  Let's assume the initial inventory at the beginning of the day,  $OO_{i,t}$  which consists of ordered but not yet delivered stock. The main experiment uses a sales loss accounting system, where unmet demand is not carried over to the next period; therefore, the inventory location is :

$$IP_{i,t} = OH_{i,t} + OO_{i,t}.$$

Before each evaluation window begins, all three strategies use the same inventory initialization to reduce the impact of initial state differences on strategy comparison:

$$OH_{i,t_0} = Q_{0.90}(D_{i,t_0}^P), \quad OO_{i,t_0} = 0.$$

Each strategy operates under the same initialization conditions for a 14-day warm-up period. This warm-up period is only used to stabilize inventory levels and is not included in performance evaluation.

Daily events are executed in a fixed sequence: the system receives overdue orders at the beginning of the day, then completes forecasting and replenishment based on the available historical information at that time; demand subsequently occurs and inventory is updated. When  $L=0$ , the order arrives on the same day; when  $L>0$ , the order arrives at the beginning of day  $t+L$ .

Let the actual available inventory before demand occurs be  $A_{i,t}$  and the sales loss be :

$$LS_{i,t} = \max(0, D_{i,t} - A_{i,t}),$$

Once the request is fulfilled, the system updates the end-of-day inventory accordingly.

$$OH_{i,t}^{end} = \max(0, A_{i,t} - D_{i,t}).$$

To ensure that the strategy comparison only reflects the differences in replenishment mechanisms, all strategies share the same demand trajectory and the same delivery rules, and the demand is not resampled during the experiment.

Inventory levels below the 0.80 quantile are defined as shortage risk; levels within the range are defined as healthy; and levels above the 0.95 quantile are defined as excess risk.

$$State_{i,t} = \begin{cases} \text{Shortage Risk,} & IP_{i,t} < Q_{0.80}(D_{i,t}^P), \\ \text{Healthy,} & Q_{0.80}(D_{i,t}^P) \leq IP_{i,t} \leq Q_{0.95}(D_{i,t}^P), \\ \text{Excess Risk,} & IP_{i,t} > Q_{0.95}(D_{i,t}^P). \end{cases}$$

To determine whether state division depends on a single threshold, this study also used 0.75–0.90, 0.75–0.95, 0.80–0.90, and 0.85–0.95 for sensitivity tests, but did not reselect the main parameters based on these results.

The main experiment compares three replenishment strategies. B1 Static follows these rules:

$$S_i = \mu_i^P + z\sigma_i^P$$

B1 selects the parameter from  $z=\{1.00, 1.28, 1.64, 1.96\}$  that is closest to the main baseline of B2 during the validation period; if there are ties, this study retains the parameter with the lower inventory. This parameter remains unchanged during the testing period.

B2 Forecast-Driven uses the exact same probabilistic demand model as Proposed, with the primary setting fixed at the 0.90 quantile as the replenishment target:

$$Q_{i,t}^{B2} = \max[0, Q_{0.90}(D_{i,t}^P) - IP_{i,t}].$$

The proposed Health-Triggered model only adds a shortage risk response to the B2 level. If :

$$IP_{i,t} < Q_{0.80}(D_{i,t}^P)$$

The system raises the target inventory level to the 0.95 quantile; other states will continue to be executed at the 0.90 quantile.

$$S_{i,t}^H = \begin{cases} Q_{0.95}(D_{i,t}^P), & IP_{i,t} < Q_{0.80}(D_{i,t}^P), \\ Q_{0.90}(D_{i,t}^P), & \text{Other states,} \end{cases}$$

$$Q_{i,t}^H = \max(0, S_{i,t}^H - IP_{i,t}).$$

When inventory is at risk of excess, both B2 and Proposed stop replenishing. Therefore, this study does not include additional disposal mechanisms such as transfer, price reduction or scrapping.

The only difference between B2 and Proposed is whether the shortage risk increases the target quantile. This setup allows the performance difference between the two strategies to more directly correspond to the incremental value of inventory status information.

RQ1 uses the protection cycle coverage failure rate to measure the coverage risk within future protection cycles:

$$CFR = \frac{1}{N} \sum_{i,t} \mathbf{1}(D_{i,t}^P > IP_{i,t}),$$

This study also calculated the average shortage exposure:

$$Shortfall_{i,t} = \max(0, D_{i,t}^P - IP_{i,t})$$

And average overexposure:

$$Excess_{i,t} = \max(0, IP_{i,t} - D_{i,t}^P)$$

RQ1 is used to test risk stratification, not to identify causal effects. Inventory status is generated only based on information available at the time of decision-making, while CFR and exposure metrics are calculated after demand is realized. The main analysis uses the B2 trajectory and is recalculated on the B1 trajectory. B1 and B2 represent two types of unhealthy feedback mechanisms, namely static and probability-driven ones; if the state rankings on the two trajectories are consistent, it can be determined whether this stratification relationship depends on a particular replenishment strategy.

RQ2 uses Fill Rate and Average Inventory as its main evaluation metrics, among which :

$$Fill\ Rate = 1 - \frac{\sum LS_{i,t}}{\sum D_{i,t}},$$

Average Inventory is the average of the ending inventory. Since the original sales volume has been proportionally transformed, inventory and exposure are reported in Normalized Units of Demand (NDU). Stockout Day Rate represents the percentage of stores—SKUs—that experienced sales losses on a daily basis, and Replenishment Day Rate represents the percentage of stores—SKUs—that received orders on a daily basis.

Service-inventory frontiers were constructed for B2 and Proposed at targets of 0.80, 0.85, and 0.90, respectively. Service matching was targeted at the B2 master baseline Fill Rate. In each Bootstrap resampling, piecewise linear interpolation was performed using adjacent target quantiles surrounding the service level on the Proposed frontier, and the Average Inventory difference at the matched service level was calculated; if the target exceeded the frontier range, no extrapolation was performed. Statistical inference was performed using 1000 store-paired clustering Bootstrap runs,

and the interpolation results were recalculated in each resampling. CV, lead time, and rolling window all reported 95% CI, with Holm correction applied to the CV test.

### 3. Research Results

#### 3.1 Prospective risk stratification validity of inventory health status

After demand recovered, the two-day cumulative demand forecast WAPE was 30.08% and WPE was 1.74%; the actual coverage rates of the 0.80, 0.90 and 0.95 quantiles were 78.3%, 88.6% and 93.5%, respectively. These results show that the forecast can form a relatively clear quantile hierarchy, but the state judgment is still affected by forecast errors.

In the B2 reference trajectory, the states of shortage risk, health and excess risk accounted for 27.8%, 59.4% and 12.8%, respectively. The specific results are shown in Table 1.

*Table 1. Prospective risk stratification validity and threshold sensitivity of SKU inventory status*

*Panel A: Main health range (0.80-0.95)*

| Inventory status | Sample percentage / % | CFR/% | Average shortage exposure/NDU | Average Overexposure /NDU |
|------------------|-----------------------|-------|-------------------------------|---------------------------|
| Shortage risk    | 27.8                  | 20.1  | 0.88                          | 0.18                      |
| healthy          | 59.4                  | 9.8   | 0.32                          | 1.14                      |
| Excess risk      | 12.8                  | 6.9   | 0.20                          | 4.41                      |

*Panel B: Sensitivity of the Healthy Zone*

| The Lower World | Upper Realm | Shortage status CFR/% | Shortage Status Average Shortage Exposure/NDU | Excessive State Average Excess Exposure/NDU |
|-----------------|-------------|-----------------------|-----------------------------------------------|---------------------------------------------|
| 0.75            | 0.90        | 22.7                  | 0.96                                          | 3.68                                        |
| 0.75            | 0.95        | 22.7                  | 0.96                                          | 4.41                                        |
| 0.80            | 0.95        | 20.1                  | 0.88                                          | 4.41                                        |
| 0.80            | 0.90        | 20.1                  | 0.88                                          | 3.68                                        |
| 0.85            | 0.95        | 17.6                  | 0.77                                          | 4.41                                        |

Under trajectory B2, the CFR and shortage exposure were higher in the shortage-risk state than in the healthy state, while the average excess exposure in the excess-risk state reached 4.41 NDUs. Under trajectory B1, the CFRs for the three states were 20.8%, 10.3%, and 7.1%, respectively; the shortage exposures were 0.91, 0.34, and 0.21 NDUs, respectively; and the excess exposure was 4.25 NDUs. Recalculation in B1 still could not eliminate the structural association resulting from the simultaneous inclusion of IP in state definition and ex-post evaluation, but the two trajectories without health feedback maintained the same risk ranking, indicating that this relationship was not unique to any particular baseline strategy. Based on this evidence, RQ1 only supports prospective risk stratification validity for inventory status, and this study does not interpret it as an independent predictive ability.

Threshold changes affect the proportion and exposure intensity of various states, but the risk ranking between the shortage and excess sides remains unchanged. Therefore, this study interprets the sensitivity results as evidence that the state construct remains stable within a local range, rather than interpreting any particular set of thresholds as the only optimal setting.

### 3.2 The Decision Value of Health-Triggered Replenishment

The final out-of-time performance of the three strategies is shown in Table 2. These results are used to compare the actual differences between static replenishment, probability-driven replenishment, and health-triggered replenishment in the same test environment.

Table 2. Out-of-time performance of different replenishment strategies

| Strategy                  | Fill Rate/% | Average Inventory/NDU | Stockout Day Rate/% | Replenishment Day Rate/% |
|---------------------------|-------------|-----------------------|---------------------|--------------------------|
| B1 Static                 | 91.84       | 14.63                 | 17.9                | 34.7                     |
| B2 Forecast-Driven        | 92.91       | 15.00                 | 16.2                | 37.4                     |
| Proposed Health-Triggered | 93.28       | 15.24                 | 15.6                | 38.2                     |

The fill rates for B1, B2, and Proposed were 91.84%, 92.91%, and 93.28%, respectively, with average inventory of 14.63, 15.00, and 15.24 NDUs. Service levels improved after probability information was incorporated into replenishment; health triggers continued to improve service, but inventory usage also increased accordingly.

Compared to B2, the proposed Fill Rate increased by 0.37 percentage points, with a 95% CI of [0.10, 0.64]; Average Inventory increased by 0.24 NDU, or 1.60%, with a 95% CI of [0.10, 0.38]. This result indicates that the service improvement brought about by health triggering is not without cost; the strategy also incurs a higher average inventory.

Figure 1 illustrates the service-inventory frontiers for the two strategies. As the baseline target quantile increases from 0.80 to 0.90, both the Fill Rate and Average Inventory of B2 and Proposed increase accordingly. Proposed generally exhibits a higher service level and higher inventory input. Therefore, simply comparing the same target quantile is insufficient to determine whether it improves the service-inventory tradeoff; it is also necessary to compare the two frontiers at similar service levels.

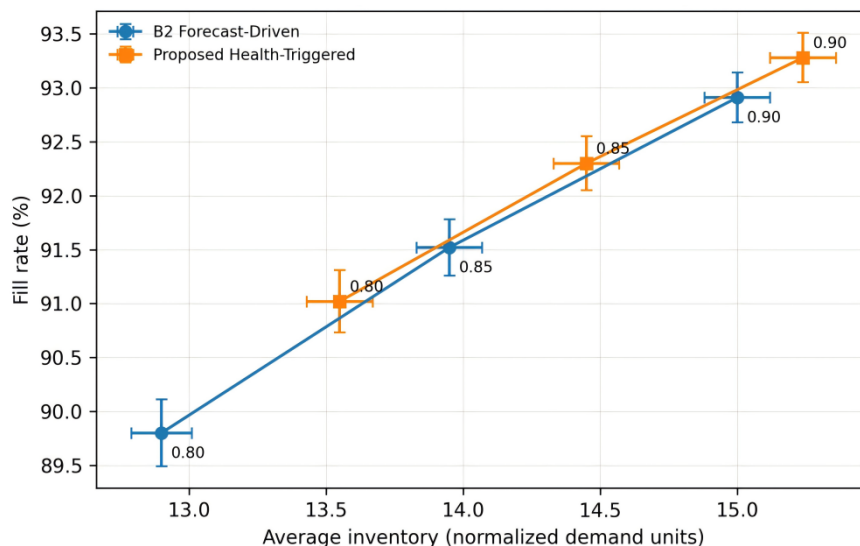


Figure 1. B2 and Proposed Services—Inventory Frontier

Note: The horizontal and vertical error bars represent the 95% confidence intervals of the store-paired clustering Bootstrap for Average Inventory and Fill Rate, respectively, with 1000 Bootstrap iterations.

All six strategy points reported a 95% CI. At the 0.90 target, the Fill Rate ranges for B2 and Proposed were [92.68%, 93.14%] and [93.05%, 93.51%], respectively, while the NDU ranges were [14.88%, 15.12] and [15.12%, 15.36]. The two ranges still have close or overlapping portions, therefore the service-inventory frontier has not achieved global separation.

After matching the service to a fill rate of approximately 92.91% using piecewise linear interpolation as described in Section 2.2, the proposed inventory was approximately 14.94 NDU, 0.06 NDU lower than B2, with a 95% CI of  $[-0.13, 0.01]$ . The upper bound of the confidence interval is close to zero but still contains zero values; therefore, this study only considers this result as directional evidence of potential inventory savings and does not confirm a stable inventory advantage.

### 3.3 Heterogeneity and Robustness

The results for CV, lead time, demand recovery, and rolling window are listed in Table 3. This study focuses on whether the incremental effects of health triggers remain in the same direction after these conditions change.

*Table 3 Heterogeneity and Robustness Tests*

| test               | Context   | B2 Fill Rate/% | Proposed Fill Rate/% | Increment/pp | 95% CI/pp       |
|--------------------|-----------|----------------|----------------------|--------------|-----------------|
| CV                 | Low       | 96.18          | 96.22                | +0.04        | $[-0.08, 0.16]$ |
| CV                 | Medium    | 93.07          | 93.31                | +0.24        | $[-0.05, 0.53]$ |
| CV                 | High      | 89.32          | 90.03                | +0.71        | $[0.22, 1.20]$  |
| Lead time          | L=0       | 95.32          | 95.39                | +0.07        | $[-0.05, 0.19]$ |
| Lead time          | L=1       | 92.91          | 93.28                | +0.37        | $[0.10, 0.64]$  |
| Lead time          | L=2       | 89.82          | 90.36                | +0.54        | $[0.19, 0.89]$  |
| No need to restore | L=1       | 92.70          | 92.97                | +0.27        | $[0.02, 0.52]$  |
| Scrolling window   | Window 1  | 92.44          | 92.63                | +0.19        | $[-0.03, 0.41]$ |
| Scrolling window   | Windows 2 | 91.76          | 92.17                | +0.41        | $[0.08, 0.74]$  |
| Scrolling window   | Windows 3 | 93.18          | 93.48                | +0.30        | $[0.02, 0.58]$  |

The fill rate increases for the low, medium, and high CV groups were 0.04, 0.24, and 0.71 percentage points, respectively. The 95% CI for the low and medium CV groups crossed zero, while the 95% CI for the high CV group was  $[0.22, 1.20]$ , with a Holm-corrected  $p=0.008$ , and inventory increased by 1.9%. Looking at the changes between groups, the greater the demand volatility, the more significant the healthy-triggered service increase, but this improvement was still accompanied by inventory input.

When  $L=0, 1$ , and  $2$ , the Fill Rate increments are 0.07, 0.37, and 0.54 percentage points, respectively, with 95% CIs of  $[-0.05, 0.19]$ ,  $[0.10, 0.64]$ , and  $[0.19, 0.89]$ . The increment increases with the length of the lead time, but this study only interprets this phenomenon as a trend within the applicable boundary and does not further infer a linear relationship.

After the demand recovery was cancelled, the fill rates for B2 and Proposed were 92.70% and 92.97%, respectively, with inventories of 14.86 and 15.09 NDU, and stockout day rates of 16.6% and 16.1%, respectively. Compared to the main experiment, the main change came from the underestimation of potential demand, while the direction of improvement of the healthy trigger relative to B2 remained unchanged.

The fill rate increments for the three rolling windows were 0.19, 0.41, and 0.30 percentage points, respectively, with 95% CIs of  $[-0.03, 0.41]$ ,  $[0.08, 0.74]$ , and  $[0.02, 0.58]$ . The interval of Window 1 crossed zero, indicating that the increment under the short window was not stable; the latter two windows maintained positive results, providing supplementary evidence for the main conclusion.

## 4. Discussion

### 4.1 Key Findings and Management Implications

The findings indicate that the primary function of inventory health is not to improve forecast accuracy, but rather to transform existing probabilistic demand information into state-dependent risk control. The consistent risk ranking of both the B1 and B2 reference trajectories suggests that inventory status has a certain forward-looking stratification significance. However, this evidence still comes from a post-hoc comparison of inventory location with actual demand; therefore, this study does not interpret it as causal prediction. Probabilistic prediction answers the question "How high is the future demand?", while inventory health further answers "Where does the current inventory position within the future demand distribution?"

This design examines the value of two types of information separately. B1 to B2 primarily reflects the effect of probabilistic demand entering replenishment, while B2 to Proposed adds state feedback under the condition that the forecast remains unchanged, corresponding to the marginal value of inventory state information. Since B2 already incorporates demand quantiles into the ordering target, Proposed only increases the fill rate by 0.37 percentage points, and the inventory difference after service matching is also at the statistical boundary. This indicates that the effect of the added state information has significant diminishing marginal returns. In other words, the difference between this study and classic state-dependent inventory control lies not in the action of "increasing inventory after falling below a threshold" itself, but in the fact that the threshold is dynamically defined by the distribution of forward demand, and the additional effect brought by state information is identified separately through a strong baseline.

The more pronounced increments under high CV and longer lead times provide further clues to the mechanism. One possible reason is that with increased demand volatility or extended protection periods, the uncertainty of cumulative demand is more strongly exposed, making it easier for inventory levels to fall below the 0.80 quantile when the target is fixed at 0.90. The healthy trigger, by raising the target to 0.95, effectively adds a buffer against tail shortage risks. This explanation aligns with the heterogeneity results, but this study did not examine the mediating relationship between trigger frequency, quantile spacing, and incremental performance; therefore, this judgment remains a mechanistic explanation rather than a causal proof.

The results also bring a noteworthy management implication: more status information does not automatically translate into lower inventory. The proposed solution, while improving service levels at a target of 0.90, increased average inventory by 1.60%; and after service matching, there was no stable evidence of inventory savings. For practical replenishment management, this means that companies should not uniformly increase risk targets for all SKUs. Health triggers are more suitable for SKUs with high CV and long lead times, and their application requires comparing trigger frequency, inventory increase, and service benefits simultaneously.

Existing DRL methods can handle inventory optimization under non-stationary demand, and Deep Controlled Learning is also used for learning-based inventory control. This study differs from these methods in its research focus and does not pursue the absolute advantage of black-box strategies on a single indicator. Instead, it retains a verifiable decision path: strategy changes can be traced back to inventory location, protection cycle prediction quantile, and target adjustment. This design facilitates explanation of why replenishment is triggered and allows companies to adjust thresholds based on business rules and risk preferences. Therefore, the main contribution of this study is closer to interpretable risk calibration than to proposing a new control mechanism.

## 4.2 Limitations and Future Research

This study is still limited by data availability. FreshRetailNet-50K does not publicly provide complete daily inventory, purchase orders, and delivery logs. Therefore, this study can only reconstruct inventory evolution based on actual demand trajectories and cannot directly interpret the simulation results as actual operational improvements after implementation by the original enterprise. Furthermore, the original sales volume has undergone a uniform proportional transformation, so NDU can only be used for relative comparisons between strategies and cannot be restored to the actual number of items.

Both the state definition and post-evaluation of RQ1 include IP. Even if the trajectories of B1 and B2 are consistently ordered, the association introduced by this definitional structure cannot be completely eliminated. The 95% CI for the inventory difference after service matching is  $[-0.13, 0.01]$ . This study continues to use the pre-set 95% confidence level and does not redefine significance by lowering the confidence level. The official final testing period is only 7 days. Although the rolling window can supplement evidence of time stability, it still cannot replace longer-term external validation.

This study has not yet addressed shelf life, random lead time, order quantity, multi-warehouse coordination, and supply disruptions. Demand recovery errors and quantile forecast errors will continue to propagate to inventory decisions. Future research requires ERP or WMS data that includes actual inventory, ordering, and delivery records, and will examine the external stability of inventory health over longer periods and in different product environments. The current health status only reflects the risk position at the decision-making point. Future research could further model the state transition probability, extending the "health-shortage-oversupply" stratification from the current period to a predictable risk change process, and testing whether this predictive health status can intervene in advance without increasing average inventory.

## 5. Conclusion

This study distinguishes the incremental decision-making value of probabilistic demand information and inventory status information. The study defines a two-sided risk status based on the relationship between SKU inventory location and protection cycle demand distribution, and examines the actual role of status information under the same forecasting environment through strong baseline comparison. The results show that inventory health status can provide forward-looking risk stratification, but this information does not universally improve service-inventory efficiency; its incremental value mainly occurs in high-uncertainty scenarios such as high volatility and long lead times. Therefore, inventory health is more suitable as an interpretable and auditable risk calibration interface connecting probabilistic forecasting and replenishment rules, rather than a uniform optimization standard applicable to all SKUs.

## References

- [1] Trapero J R, Holgado de Frutos E, Pedregal D J. Demand forecasting under lost sales stock policies[J]. *International Journal of Forecasting*, 2024, 40(3): 1055-1068. DOI: 10.1016/j.ijforecast.2023.09.004.
- [2] Theodorou E, Spiliotis E, Assimakopoulos V. Forecast accuracy and inventory performance: Insights on their relationship from the M5 competition data[J]. *European Journal of Operational Research*, 2025, 322(2): 414-426. DOI: 10.1016/j.ejor.2024.12.033.

- [3] Hasni M, Babai M Z, Rostami-Tabar B. A hybrid LSTM method for forecasting demands of medical items in humanitarian operations[J]. *International Journal of Production Research*, 2024, 62(17): 6046-6063. DOI: 10.1080/00207543.2024.2306904.
- [4] Wang S, Kang Y, Petropoulos F. Combining probabilistic forecasts of intermittent demand[J]. *European Journal of Operational Research*, 2024, 315(3): 1038-1048. DOI: 10.1016/j.ejor.2024.01.032.
- [5] Van der Haar J F, Wellens A P, Boute R N, Basten R J I. Supervised learning for integrated forecasting and inventory control[J]. *European Journal of Operational Research*, 2024, 319(2): 573-586. DOI: 10.1016/j.ejor.2024.07.004.
- [6] Wang, B. (2025). *Methods of Load Optimization for Computer Systems Based on Physical Principles*.
- [7] Guo, X. (2025, March). *Research on Blockchain-Based Financial AI Algorithm Integration Methods and Systems*. In *2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE)* pp.816 – 821. IEEE.
- [8] Guo, X. (2025, April). *Research on Financial Trading Algorithms Based on Deep Reinforcement Learning*. In *2025 IEEE 3rd International Conference on Control, Electronics and Computer Technology (ICCECT)* pp.1711 – 1715. IEEE.
- [9] Yu, X. (2026). *Exploration of Multi-Channel Conversion Path Optimization Methods Based on A/B Testing*.
- [10] Jing, X. (2025, November). *In-Depth Analysis of the Optimization Decision Model of Financial Investment Portfolio Based on Machine Learning Algorithms in Python Scikit-Learn Library*. In *International Conference on Computational Technologies for Research in Data Analytics* pp.130 – 140. Cham: Springer Nature Switzerland.
- [11] Liu, X. (2026). *Research on the Application of Artificial Intelligence in Consumer Privacy Data Protection*. *Procedia Computer Science*, 279, 956-965.
- [12] Liu, X. (2026). *Construction of Consumer Data Privacy Protection System Based On Blockchain Technology*. *Procedia Computer Science*, 282, 2082-2091.
- [13] Huijie Pan. *Discussion on Low-Latency Computing Strategies in Real-Time Hardware Generation*. *International Journal of Neural Network* (2025), Vol. 4, Issue 1: 57-64.
- [14] Han, W. (2025, November). *An Empirical Study on Building a High-precision Forecasting System for Financial Market Trends Based on the PyTorch Deep Learning Framework*. In *International Conference on Computational Technologies for Research in Data Analytics* (pp. 518-528). Cham: Springer Nature Switzerland.
- [15] Yuan, Y. (2026). *Research on Memory Management and Dynamic Reasoning Methods for Intelligent Agents Based on Large Language Models*.
- [16] Yang, Y. (2026). *Causal Inference-based Identification of Incremental Effects in Digital Advertising and Optimization Pathways for Resource Allocation*. *Advances in Computer and Communication*, 7(2).
- [17] Sun, L. (2026). *Consistency Optimization Strategies for Distributed Advertising Systems Under Cross-regional Deployment*. *Advances in Computer and Communication*, 7(2).