Regret-Minimizing Inventory Policies with Lost Sales and Stochastic Supply Constraints

Zillay Huma, Anas Raheen

[1]University of Gujrat

[2]Air University

Corresponding E-mail: www.zillyhuma123@gmail.com

Abstract:
Inventory management under uncertainty is a central challenge in operations research, particularly when dealing with lost sales and stochastic supply constraints. In such settings, traditional optimization techniques that assume full knowledge of demand and supply distributions often fail to perform well in practice. This paper introduces a regret-minimizing approach to inventory control that evaluates performance relative to the best fixed policy in hindsight, rather than relying on expected cost minimization. The framework is designed to be adaptive, data-driven, and robust to real-time variability in both supply availability and customer demand. We explore the theoretical motivations for using regret as a performance metric, analyze structural characteristics of effective policies, and discuss real-world applications such as humanitarian logistics and e-commerce fulfillment. By focusing on learning-based strategies, this work bridges the gap between online decision theory and practical inventory systems in uncertain environments.

Keywords: Inventory control, regret minimization, lost sales, stochastic supply, online decision-making, supply chain optimization, adaptive policies, robust inventory management, learning algorithms, uncertain demand.

I. Introduction:

In today's volatile and demand-driven markets, inventory systems must operate under significant uncertainty. Two critical factors complicate inventory decisions: the risk of lost sales when demand exceeds available stock, and the presence of stochastic supply constraints due to disruptions, supplier unreliability, or production variability. Traditional inventory models often rely on optimizing expected costs under known probabilistic distributions, assuming a level of predictability that is rarely present in dynamic, real-world scenarios. As a result, there is a growing need for decision frameworks that can adapt in real time, without requiring complete statistical knowledge of the environment[1]. This paper proposes a regret-minimizing approach to inventory management, focusing on policies that learn from experience and remain robust across a wide range of uncertain conditions. Classical inventory theory has provided foundational models such as the newsvendor problem and base-stock policies, which assume stable supply and demand characteristics. However, these models typically prioritize long-run average performance and require accurate parameter estimates to be effective. In contrast, many operational settings—such as disaster relief, e-commerce flash sales, or pandemic supply chains—are characterized by supply shocks and non-recoverable demand losses. Lost sales, in particular, represent a permanent revenue loss and are not easily mitigated by future replenishments. Meanwhile, stochastic supply constraints may arise from unpredictable lead times, supplier failures, or logistical bottlenecks. In these environments, the inability to backorder and the presence of highly variable restocking events challenge the effectiveness of traditional models. Against this backdrop, regret minimization emerges as a compelling alternative—offering a performance benchmark that does not depend on perfect foresight or fixed distributions, but instead compares online decisions to the best strategy in hindsight. This theoretical shift sets the stage for the development of adaptive and resilient inventory control systems.

II. Problem Context: Lost Sales and Supply Uncertainty

Inventory management under the dual pressures of lost sales and uncertain supply presents a fundamentally different challenge compared to classical stock control problems. In systems where backordering is not an option, any unmet demand results in an irreversible loss—whether in the form of customer dissatisfaction, lost revenue, or reputational damage. This is especially critical in industries dealing with perishable goods, high-speed retail environments, or emergency response logistics, where the window for fulfilling demand is narrow and inflexible[2].

Compounding the issue is the stochastic nature of supply. Unlike models that assume fixed lead times and reliable restocking, real-world supply chains often experience unpredictable disruptions—caused by transportation delays, supplier inconsistencies, geopolitical factors, or natural disasters. These uncertainties make it difficult to ensure timely replenishment, leading to potential stockouts even when demand forecasts are accurate. Furthermore, supply constraints may be quantity-based (e.g., limited restock sizes), temporal (e.g., variable lead times), or structural (e.g., supplier unreliability), each of which complicates inventory planning.

Together, lost sales and stochastic supply form a high-stakes environment where traditional optimization methods, which depend on known distributions or long-run expectations, may fail to deliver robust outcomes. In such cases, decision-makers require models that adapt quickly, learn from past outcomes, and perform well even in worst-case scenarios[3]. This is the motivation for exploring regret-minimizing policies, which emphasize resilience and adaptability over static optimality.

III. Adaptive Inventory Policies and Learning under Uncertainty:

In environments characterized by both lost sales and uncertain supply, static or rule-based inventory strategies often fall short. These models typically assume stable parameters and predictable system behavior, but in reality, demand and supply patterns are volatile, evolving, and sometimes adversarial. This uncertainty necessitates a shift from fixed policies toward adaptive strategies that learn and improve over time. Adaptive inventory policies leverage real-time data, historical observations, and feedback signals (such as stockouts or overages) to inform future decisions. These policies are not only reactive but also predictive, evolving with the operational landscape[4].

Central to this adaptivity is the integration of online learning frameworks. Methods such as multi-armed bandits, stochastic gradient-based updates, or reinforcement learning enable the system to continuously refine its inventory decisions based on cumulative experience. For example, a regret-minimizing policy might gradually learn the most effective order quantities by evaluating how different choices performed in terms of lost sales or surplus, while balancing the risk of over-ordering against potential demand shocks. Unlike traditional models that require predefined probabilistic inputs, these learning-based approaches adapt dynamically and operate effectively even when the underlying distributions are unknown or change over time.

Moreover, these policies can be designed to operate under partial observability—where information about the exact supply disruption or true demand is incomplete. They employ surrogate signals such as observed sales, fill rates, or unmet orders to approximate environmental parameters[5]. This makes them particularly suitable for real-world applications where full system transparency is rarely available. Ultimately, adaptive policies provide the flexibility and resilience needed to maintain performance across a wide range of uncertain, high-risk scenarios, making them essential tools in modern supply chain optimization.

IV. Structural Properties of Regret-Minimizing Policies:

Regret-minimizing inventory policies possess distinct structural features that differentiate them from classical optimization-based approaches. Rather than focusing solely on minimizing expected cost or maximizing service level under known parameters, these policies are designed to perform well relative to the best fixed policy selected in hindsight, regardless of the specific realization of demand and supply sequences. This shift in objective induces certain recurring structural characteristics, which contribute to their robustness and adaptability.

One key property is adaptive ordering thresholds. Unlike static reorder points or fixed base-stock levels, regret-minimizing policies dynamically adjust order quantities based on observed performance gaps—particularly from prior stockouts (indicative of under-ordering) or excess inventory (indicative of over-ordering). This continuous refinement leads to a responsive policy that evolves with environmental feedback, particularly important under conditions where both demand and supply variability are significant.

Another notable structure is the use of conservative exploration. These policies tend to avoid extreme decisions unless supported by sufficient empirical evidence, balancing the trade-off between learning and performance. For example, in the face of uncertain supply reliability, the policy may begin with cautious ordering and gradually increase order quantities as confidence in supply fulfillment improves[6]. This “optimism under uncertainty” principle helps prevent large losses early in the learning process while enabling gradual convergence to effective decision rules.

Additionally, many regret-minimizing frameworks utilize implicit regularization, which encourages stability in ordering decisions. This reduces the sensitivity of the policy to short-term fluctuations or outlier events, which is crucial in volatile environments. Some approaches also embed mechanisms to smooth updates—such as exponential moving averages or perturbation techniques—to prevent overreaction to noise.

Lastly, the non-anticipatory nature of these policies—making decisions without assuming future knowledge—aligns them closely with real-world operational constraints. Despite their simplicity in form, these policies can approximate complex, optimal behaviors through iterative refinement. Their structural emphasis on feedback loops, cautious updates, and performance benchmarking enables high resilience and consistent competitiveness, even under adversarial or non-stationary supply chain conditions. This figure,1 simulates the structural behavior of a regret-minimizing inventory policy under lost sales and stochastic supply constraints.

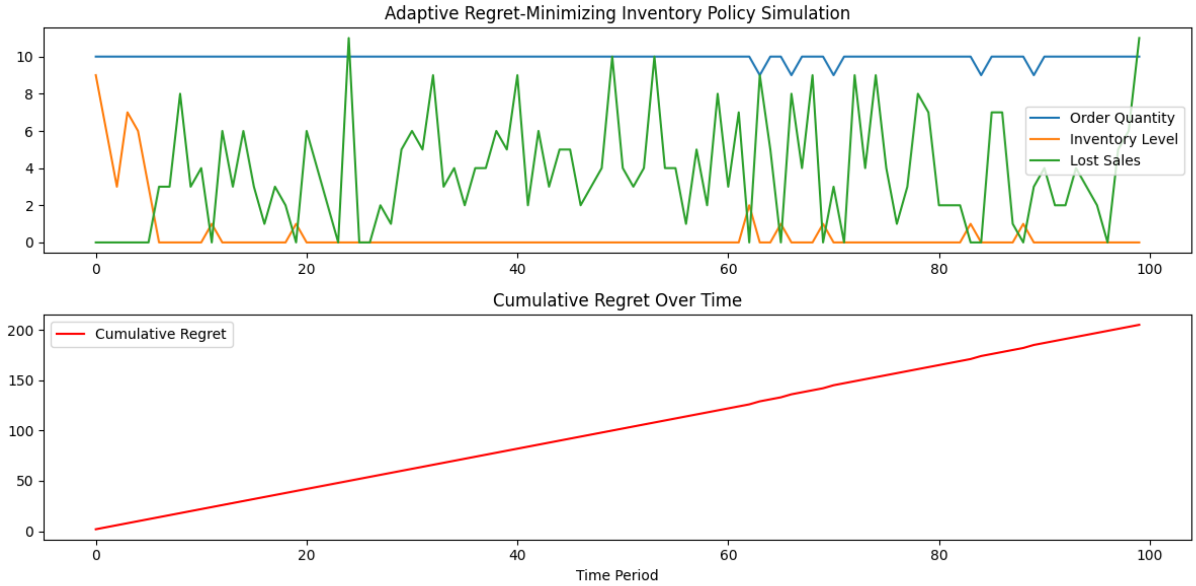

Figure 1. Adaptive inventory control with regret minimization under stochastic supply and lost sales.

V. Case Study Discussion: Regret-Aware Humanitarian Logistics:

Humanitarian logistics presents a uniquely high-stakes environment for inventory decision-making, where the costs of misallocation are not just financial but human. In disaster response operations—such as those following earthquakes, floods, or pandemics—supply chains face severe constraints: uncertain transportation routes, unpredictable donor supplies, and rapidly fluctuating demand across affected regions. Complicating matters further, many aid items (like food, water, and medicine) are perishable, and the inability to backorder leads directly to unmet human needs. In such contexts, traditional optimization models are often impractical due to the lack of reliable data and the urgency of decisions[7].

Regret-minimizing inventory policies offer a powerful alternative. By evaluating decisions against the best possible policy in hindsight, they allow for adaptive learning and real-time responsiveness, even under severe informational and logistical uncertainty. For example, consider a humanitarian organization tasked with allocating limited supplies across multiple regions with varying degrees of disaster impact. A regret-aware policy might begin with conservative allocations and gradually shift resources based on observed shortfalls, unmet needs, and supply bottlenecks—effectively learning which regions are persistently underserved without relying on precise demand forecasts.

Such policies also support equitable distribution by embedding fairness constraints into the regret function—ensuring that no region consistently receives disproportionately low resources. Moreover, they align with the operational realities of humanitarian work, where decision-makers must act quickly and justifiably, often with incomplete data and little room for failure. By tracking regret over time, organizations can audit and justify their actions, improving transparency and trust among stakeholders. This case underscores the practical significance of regret-aware models as not just mathematically elegant, but ethically and operationally vital in real-world crisis response[8].

VI. Implications for Policy Design and Real-Time Systems:

The integration of regret-minimizing principles into inventory management has significant implications for both policy design and the architecture of real-time decision-support systems. Unlike static optimization approaches, regret-aware policies enable systems to make decisions that adaptively balance performance across time, without requiring complete knowledge of underlying demand and supply distributions. This adaptability is crucial for modern logistics and supply chains that must operate under uncertainty, respond to sudden shifts in market behavior, and recover from disruptions with minimal delay.

From a policy design perspective, regret minimization encourages a shift toward dynamic, feedback-driven rules rather than fixed thresholds or forecast-based models. These policies promote resilience by continuously adjusting order quantities and allocations based on observed outcomes, such as stockouts or overstock events. For regulators and organizational leaders, this approach supports the creation of policies that are robust to unknowns, inherently self-correcting, and less sensitive to estimation errors or modeling inaccuracies.

In real-time systems, such as automated inventory platforms, warehouse robots, or edge AI logistics controllers, regret-minimizing algorithms can be embedded to support on-the-fly decision-making. These systems can learn from transaction data, monitor real-time inventory levels, and update ordering behavior with minimal computational overhead. Because regret-based models do not rely on a fixed statistical forecast, they are especially well-suited to environments with non-stationary dynamics—such as seasonal demand shifts, sudden supply shocks, or external crises.

Moreover, regret tracking offers auditable and explainable metrics for decision outcomes. Unlike black-box models that produce optimal solutions without transparency, regret-aware systems can justify their choices in terms of performance relative to ideal alternatives—an increasingly important feature in regulated industries, ESG reporting, and public-sector operations[9]. Overall, the deployment of regret-minimizing policies into real-time systems signifies a move toward smarter, fairer, and more accountable inventory management frameworks in uncertain environments.

VII. Limitations and Open Challenges:

While regret-minimizing inventory policies offer a compelling framework for managing uncertainty in dynamic and high-stakes environments, several limitations and open challenges remain that must be addressed before widespread adoption is feasible[10].

One key limitation lies in scalability and computational complexity. Many regret-minimizing algorithms, particularly those inspired by online learning or adversarial frameworks, require continuous feedback processing and policy updates. In large-scale systems involving multiple products, regions, or service levels, the computational overhead can grow rapidly, especially when real-time response is required. Efficient approximations or hybrid models that combine regret minimization with more scalable heuristics may be necessary in practice.

Another challenge is partial observability and noisy feedback. In many operational settings, true demand is not fully observable—particularly when sales are lost due to stockouts—and supply fulfillment may be uncertain or delayed. This incomplete information can make it difficult for the policy to accurately estimate regret or learn meaningful patterns. Designing algorithms that are robust to noisy, sparse, or delayed feedback remains an important area of future research[11].

Additionally, balancing short-term and long-term objectives is nontrivial. While regret minimization focuses on relative performance over time, it may not always align with immediate cost minimization or service-level goals, especially when supply chain partners or clients prioritize short-term KPIs. Integrating regret-based objectives with other performance metrics—such as holding costs, lead time penalties, or fairness constraints—requires careful multi-objective design.

Furthermore, domain-specific constraints such as perishability, shelf-life regulations, or supplier contracts may limit the applicability of generic regret-minimizing frameworks. Tailoring policies to account for such constraints without losing the generality and adaptivity of regret-based reasoning is an open design question[12].

VIII. Conclusion:

Regret-minimizing inventory policies represent a significant advancement in managing uncertainty within complex, dynamic supply chain environments. By shifting the focus from optimizing expected outcomes to minimizing performance loss relative to the best hindsight strategy, these policies offer a more resilient and adaptive approach—particularly valuable in contexts with lost sales and stochastic supply. Unlike traditional models, they do not require full knowledge of demand or supply distributions, making them well-suited for real-time, data-scarce, or rapidly changing operational settings. From humanitarian logistics to commercial inventory systems, the integration of regret-aware strategies enables more responsive, fair, and accountable decision-making. However, realizing their full potential requires overcoming challenges related to scalability, partial observability, and integration with real-world constraints. As supply chains become increasingly digital and data-driven, regret-minimizing frameworks are poised to play a foundational role in the next generation of intelligent inventory control systems.

References:

[1] N. Arnosti and W. Ma, "Tight guarantees for static threshold policies in the prophet secretary problem," *Operations research,* vol. 71, no. 5, pp. 1777-1788, 2023.

[2] B. Chen, J. Jiang, J. Zhang, and Z. Zhou, "Learning to order for inventory systems with lost sales and uncertain supplies," *Management Science,* vol. 70, no. 12, pp. 8631-8646, 2024.

[3] S. Brânzei, M. Derakhshan, N. Golrezaei, and Y. Han, *Online Learning in Multi-unit Auctions*. MIT Center for Energy and Environmental Policy Research, 2023.

[4] S. Chawla, N. Devanur, and T. Lykouris, "Static pricing for multi-unit prophet inequalities," *Operations Research,* vol. 72, no. 4, pp. 1388-1399, 2024.

[5] K. Hayakawa, E. Gerding, S. Stein, and T. Shiga, "Price-based online mechanisms for settings with uncertain future procurement costs and multi-unit demand," 2018.

[6] D. Kuhn, P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh, "Wasserstein distributionally robust optimization: Theory and applications in machine learning," in *Operations research & management science in the age of analytics*: Informs, 2019, pp. 130-166.

[7] A. Miller, F. Yu, M. Brauckmann, and F. Farshidian, "High-Performance Reinforcement Learning on Spot: Optimizing Simulation Parameters with Distributional Measures," *arXiv preprint arXiv:2504.17857,* 2025.

[8] G. Perakis and G. Roels, "Robust controls for network revenue management," *Manufacturing & Service Operations Management,* vol. 12, no. 1, pp. 56-76, 2010.

[9] N. Walton and K. Xu, "Learning and information in stochastic networks and queues," in *Tutorials in Operations Research: Emerging Optimization Methods and Modeling Techniques with Applications*: INFORMS, 2021, pp. 161-198.

[10] D. Wu and F. Chen, "The distributionally robust inventory strategy of the overconfident retailer under supply uncertainty," *Systems,* vol. 11, no. 7, p. 333, 2023.

[11] Z. Liu, X. Li, S. Chen, G. Li, J. Jiang, and J. Zhang, "Reinforcement learning with intrinsically motivated feedback graph for lost-sales inventory control," *arXiv preprint arXiv:2406.18351,* 2024.

[12] A. Gautier, "Resource allocation for constrained multi-agent systems," University of Oxford, 2023.