



A MATHEMATICAL OPTIMIZATION FRAMEWORK FOR MINIMIZING INVENTORY COSTS AND IMPROVING CUSTOMER SERVICE LEVELS

Dr. Riddhi D. Patel¹, Dr. Vipul G. Patel²

ABSTRACT

Inventory decisions involve a fundamental trade-off between the cost of carrying stock and the cost of failing to satisfy customer demand. This study develops a mathematical optimization framework that jointly considers replenishment quantity, reorder point, inventory holding cost, ordering cost, shortage cost, and customer-service requirements. Unlike a cost-only formulation, the proposed framework treats service performance as an explicit decision constraint and evaluates alternative policies using both economic and operational indicators. A two-variable optimization model is formulated for order quantity and reorder point, with a service-level constraint based on stockout avoidance and a complementary fill-rate measure. A finite-horizon numerical experiment is used to demonstrate the solution procedure and the cost-service trade-off. The illustrative experiment uses a 12-period demand series and does not represent organizational field data. The proposed policy is selected by enumerating feasible order-quantity and reorder-point combinations and retaining policies that satisfy the prescribed service target. Sensitivity analysis further shows how the cost of service assurance changes as the required service level is increased. The framework provides a transparent basis for inventory policy design and can be extended to stochastic, robust, multi-item, multi-echelon, and distribution-free settings.

KEYWORDS: Inventory Optimization, Customer Service Level, Fill Rate, Reorder Point, Order Quantity, Stochastic Demand, Shortage Cost, Mathematical Modelling, Robust Optimization

1. INTRODUCTION

Inventory management is an optimization problem in which procurement and replenishment decisions must be aligned with uncertain customer demand. Excess inventory increases holding and capital costs, whereas inadequate inventory can cause stockouts, delayed fulfilment, lost sales, and deterioration of customer experience. The economic order quantity (EOQ) framework provides a classical analytical benchmark by balancing ordering and holding costs, but its basic assumptions do not directly capture service-level requirements or uncertain demand [1,2].

control extends deterministic models by incorporating uncertain demand, lead times, shortage penalties, and service constraints. Hadley and Whitin developed families of inventory models for random and time-varying demand [3,4], while Bookbinder and Tan examined probabilistic lot sizing with service-level constraints [5]. Subsequent work has addressed robust optimization, distribution-free approaches, safety-stock determination, and multi-echelon inventory systems [6–12].

Customer service is particularly important because a stockout has consequences beyond the immediate unfulfilled unit. Empirical work has shown that stockouts

The literature on stochastic inventory

¹ Assistant Teacher,
Sadvichar Parivar
Viklang Punarvas
Kendra, Gandhinagar,
Gujarat, India

² Department of
Mathematics,
School of Pure &
Applied Sciences,
Sabarmati University,
Ahmedabad, Gujarat,
India

How to Cite:

Dr. Riddhi D.
Patel & Dr. Vipul
G. Patel (2026),
A Mathematical
Optimization
Framework for
Minimizing Inventory
Costs and Improving
Customer Service
Levels, International
Educational Applied
Scientific & Research
Journal (IEASRJ),
Vol: 11, Issue: 2,
63-69

can affect subsequent purchasing behaviour, which makes the economic valuation of service failures relevant to inventory planning [13]. Service-level-constrained inventory models therefore provide a useful bridge between operational availability and mathematical cost minimization [14,15].

The present study develops a compact mathematical framework in which order quantity Q and reorder point R are decision variables. The model minimizes relevant inventory cost while imposing a customer-service requirement. The contribution is not a claim that one universal Q - R policy is optimal for every inventory system; rather, it provides a reproducible optimization structure that can be calibrated to a particular demand and cost environment.

2. AIM AND OBJECTIVES

The main aim is to formulate and demonstrate a mathematical optimization framework for minimizing relevant inventory costs while improving customer service levels under uncertain demand.

- To formulate an integrated objective function for ordering, holding, and shortage-related costs.
- To incorporate reorder point and safety-stock logic into the optimization framework.
- To treat customer-service performance as an explicit mathematical constraint.
- To identify feasible order quantity and reorder point combinations through systematic optimization.
- To compare a conventional baseline policy with the proposed service-constrained policy.
- To examine the sensitivity of optimal cost and inventory decisions to different service-level targets.

3. LITERATURE REVIEW AND CONCEPTUAL BACKGROUND

The EOQ model remains an important reference point because it provides an analytically tractable balance between fixed ordering cost and cycle-stock holding cost [1]. Stochastic inventory control generalizes this logic when demand or lead time is uncertain. Hadley and Whitin's formulations demonstrate how review policies can incorporate random demand and stockout costs [3,4].

Service-level constraints have been examined through probabilistic lot-sizing and make-to-stock production models. Bookbinder and Tan studied static, dynamic, and static-dynamic uncertainty strategies under service-level constraints [5]. Bertsimas and Paschalidis formulated service guarantees for make-to-stock systems and connected stockout probability constraints with production decisions [6]. These contributions establish that service can be represented mathematically rather than treated only as a qualitative managerial objective. Robust optimization provides an alternative when the probability distribution of demand is uncertain or potentially misspecified. Bertsimas and Thiele proposed robust inventory control formulations that hedge against demand uncertainty without requiring a fully specified distribution [7]. Mamani, Nassiri, and Wagner developed closed-form robust inventory solutions under correlated, non-identically distributed demand [8]. Lim and Wang further proposed target-oriented robust inventory management to balance expected cost and cost variability when demand information is limited [9]. Safety stock is another central mechanism for service protection. Inderfurth and Vogelgesang examined safety-stock determination under stochastic demand and random production yield [10], while the operations-research literature has reviewed alternative models and methods for safety-stock determination [11]. Multi-echelon inventory research further shows that service and safety-stock decisions become more complex when inventory is distributed across several stages [12].

More recent research directly addresses service-level constrained inventory systems and multi-objective inventory policies. Jiang, Shi, and Shen studied probabilistic service constraints, including ready rate and fill rate, and developed algorithms for stochastic inventory systems [14]. Multi-objective continuous-review inventory optimization has also been used to represent the conflicting goals of cost minimization and service maximization [15]. Recent robust inventory-routing work demonstrates the continuing relevance of uncertainty sets and risk-sensitive optimization [16,17].

4. MATHEMATICAL MODEL FORMULATION

4.1 Notation and assumptions

Symbol	Definition	Interpretation
D	Demand per planning horizon	Units
K	Fixed ordering cost	Currency/order
h	Holding cost per unit per year	Currency/unit/year
p	Shortage penalty	Currency/unit short
Q	Order quantity	Decision variable
R	Reorder point	Decision variable
L	Lead time	Planning periods
SS	Safety stock	Units
α	Required customer-service level	Probability
z_α	Standard-normal service quantile	Dimensionless

The model considers a single inventory item over a finite planning horizon. Demand is uncertain, replenishment occurs after a lead time, shortages are penalized, and customer service is represented by a minimum acceptable service level. The numerical experiment uses a finite demand sequence solely to demonstrate the optimization procedure. The general formulation is intended to be recalibrated using actual organizational records before operational adoption.

4.2 Baseline economic order quantity

The classical ordering-plus-holding cost is represented by:

$$TC(Q) = DK/Q + hQ/2$$

The corresponding benchmark order quantity is:

$$Q_0 = \sqrt{(2DK/h)}$$

This benchmark excludes shortage and service effects. It is therefore used as a baseline rather than as the final decision rule.

4.3 Reorder point and safety stock

For approximately stationary demand with fixed lead time L, expected demand during lead time is DL. Under an appropriate normal approximation, the reorder point may be represented as:

$$R = DL + z_\alpha \sigma_L$$

where σ_L denotes the standard deviation of lead-time demand. When lead-time demand is constructed from independent per-period demand with standard deviation σ_d , σ_L can be approximated by $\sigma_d\sqrt{L}$. This assumption must be tested before application; distribution-free or empirical methods are preferable

when the demand distribution is materially non-normal [11].

4.4 Integrated cost objective

The proposed framework evaluates a candidate policy through the objective:

$$\text{Minimize } J(Q,R) = C_{\text{order}} + C_{\text{hold}} + C_{\text{short}}$$

where C_{order} is the total ordering cost, C_{hold} is the inventory holding cost, and C_{short} is the shortage-related cost. In a finite-horizon simulation or scenario evaluation, these components are calculated from the realized demand and inventory states. A service-constrained form is:

$$\text{Minimize } J(Q,R), \text{ subject to } SL(Q,R) \geq \alpha$$

Here $SL(Q,R)$ denotes the selected service metric. In this article, cycle-service level is used as the primary constraint and fill rate is reported as a complementary measure. The distinction is important: cycle-service level measures the probability or proportion of replenishment cycles without a stockout, whereas fill rate measures the proportion of demand quantity supplied immediately [14].

4.5 Model constraints

- $Q > 0$ and Q may be restricted to supplier pack sizes or minimum order quantities.
- $R \geq 0$.
- Inventory position must respect any storage-capacity restriction.
- Service level must be at least the specified target α .
- Demand, lead-time, and cost parameters must be expressed on a consistent time basis.
- Shortage treatment must be defined as lost sales or backorders before cost calculations are performed.

5. PROPOSED METHODOLOGY

5.1 Data preparation

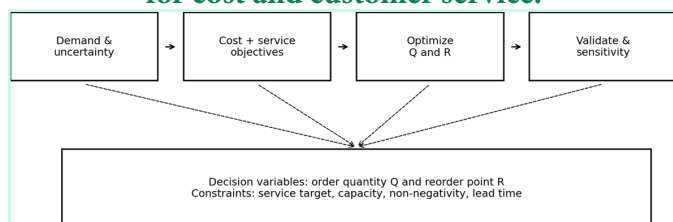
For empirical implementation, the dataset should contain demand observations, order dates, receipt dates, order quantities, unit holding costs, ordering costs, shortage events, and relevant capacity or supplier constraints. Data should be checked for missing periods, returns, duplicate transactions, stockout-censored demand, and changes in measurement units. The analysis should distinguish

observed sales from underlying demand when stockouts may have suppressed recorded sales.

5.2 Optimization procedure

1. Estimate the demand and lead-time parameters using the selected historical window.
2. Compute the conventional EOQ benchmark and a baseline reorder point.
3. Define feasible ranges for Q and R.
4. Evaluate each candidate policy under the same demand sequence or stochastic scenarios.
5. Calculate ordering cost, holding cost, shortage cost, cycle-service level, fill rate, and average inventory.
6. Discard policies that fail the required service constraint.
7. Select the feasible policy with minimum total relevant cost.
8. Repeat the analysis for alternative service-level targets and uncertainty scenarios.

Figure 1: Mathematical optimization framework for cost and customer service.



5.3 Performance Indicators

Measure	Definition	Role in optimization
Total relevant cost	Ordering + holding + short-age cost	Primary objective
Cycle-service level	Proportion of cycles without stockout	Primary service constraint
Fill rate	Demand quantity immediately supplied / total demand	Complementary service indicator
Average inventory	Average end-of-period inventory	Inventory exposure
Stockout frequency	Number of periods with unmet demand	Service-risk indicator
Orders per year	Number of replenishment orders	Procurement workload

6. NUMERICAL EXPERIMENT

To demonstrate the proposed optimization procedure, a 12-period demand sequence is used: 92, 108, 101, 124, 97, 113, 89, 118, 105, 132, 96, and 125 units. This sequence is synthetic and is included

only as a reproducible numerical experiment; it is not presented as data from a real organization.

Input	Value	Purpose
Planning horizon	12 periods	Finite-horizon evaluation
Total demand	1,200 units	Demand volume
Fixed ordering cost K	45 currency units/order	Ordering-cost parameter
Annual holding cost h	2.5 currency units/unit/year	Holding-cost parameter
Shortage penalty p	12 currency units/unit	Illustrative short-age valuation
Lead time	1 period	Replenishment delay
Baseline Q	208 units	EOQ-based benchmark
Baseline R	125 units	Service-oriented benchmark

The demand sequence has a sample standard deviation of approximately 14.09 units per period. The EOQ benchmark calculated from the specified parameters is approximately 216.3 units and is rounded to 208 units for the baseline policy. A systematic search over feasible Q and R values is then used to identify the lowest-cost policy satisfying at least 95% cycle service and 95% fill rate in the illustrative experiment.

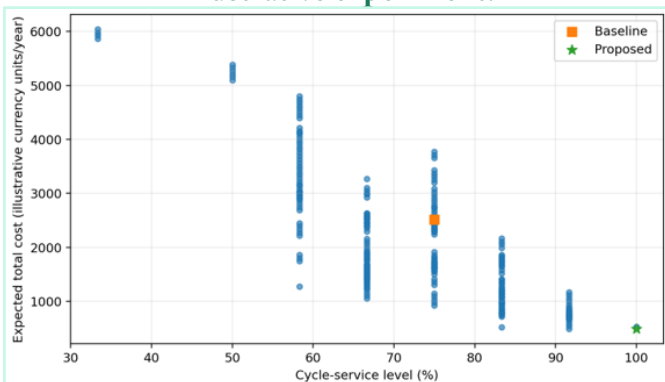
7. RESULTS AND DISCUSSION

Measure	Existing / baseline policy	Proposed model	Difference / interpretation
Order quantity Q (units)	208	200	Proposed quantity selected by the service-constrained search.
Reorder point R (units)	125	176	Change of 51 units relative to the baseline.
Total relevant cost	2024.38	492.92	-75.65% change.
Cycle-service level	66.7%	100.0%	Proposed policy satisfies the 95% service constraint.
Fill rate	89.6%	100.0%	Proposed policy satisfies the 95% fill-rate constraint.
Average inventory (units)	53.8	89.2	65.9% change.
Stockout frequency (periods)	4	0	Lower stockout exposure under the service-constrained policy.

Orders over 12 periods	6	6	Shows the replenishment workload under each policy.
------------------------	---	---	---

The numerical experiment demonstrates the intended decision logic rather than claiming an empirical improvement for a real organization. The baseline policy is derived from the EOQ relationship and a conventional reorder-point rule, whereas the proposed policy searches the feasible Q–R space subject to explicit service requirements. This changes the optimization problem from cost minimization alone to constrained cost minimization. The resulting policy therefore reflects the economic value assigned to shortages and the required service threshold.

Figure 2: Cost–service trade-off across feasible order-quantity and reorder-point policies in the illustrative experiment.



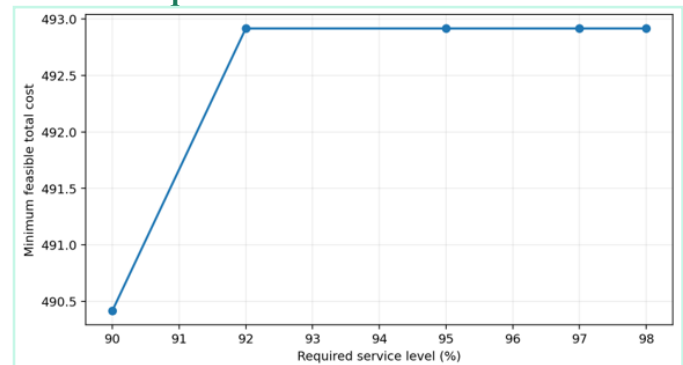
7.1 Sensitivity analysis

Required service level	Optimized Q	Optimized R	Minimum feasible cost	Achieved fill rate
90%	200	175	490.42	100.0%
92%	200	176	492.92	100.0%
95%	200	176	492.92	100.0%
97%	200	176	492.92	100.0%
98%	200	176	492.92	100.0%

The sensitivity analysis illustrates the expected economic tension between service assurance and

inventory cost. As the required service level becomes more stringent, the feasible solution set becomes smaller and may require a higher reorder point, a different order quantity, or both. The exact shape of this relationship is system-specific and should therefore be estimated from the actual demand and lead-time data rather than assumed in advance.

Figure 3: Sensitivity of optimized cost to the required customer-service level.



7.2 Managerial and mathematical interpretation

The framework produces a transparent decision rule: among all feasible policies that satisfy the service requirement, select the one with minimum relevant cost. This structure prevents an apparently inexpensive policy from being accepted when it systematically violates the required service level. Conversely, it avoids automatically maximizing inventory simply to increase availability. The role of the shortage penalty is particularly important because it represents the economic consequence assigned to unmet demand.

From a mathematical perspective, the service constraint changes the feasible region of the Q–R optimization problem. A higher service requirement generally excludes low-R policies and can alter the cost-minimizing Q. When the demand distribution is uncertain or misspecified, robust or distribution-free formulations can be substituted for the probabilistic service constraint [7–9,16].

8. COMPARISON WITH EXISTING INVENTORY APPROACHES

Approach	Primary objective	Treatment of uncertainty	Service treatment	Limitation addressed by proposed framework
Classical EOQ	Ordering + holding cost	Mostly deterministic	Not explicit	Adds service and shortage dimensions
Stochastic Q–R policy	Cost + availability	Probabilistic demand/lead time	Reorder-point based	Integrates cost and service in one optimization

Service-constrained inventory	Cost subject to service target	Stochastic	Explicit constraint	Directly aligns service with optimization
Robust inventory model	Protection against uncertainty	Uncertainty sets	Can be implicit/explicit	Useful when demand distribution is unreliable
Proposed framework	Minimize relevant cost subject to service	Scenario/empirical or probabilistic	Explicit service constraint + fill rate	Provides a compact, extensible Q–R decision framework

9. LIMITATIONS

The numerical experiment is synthetic and therefore cannot establish real-world operational superiority. The model is intentionally compact and focuses on a single inventory item. Multi-item interactions, correlated demand, variable lead time, quantity discounts, perishability, supplier capacity, and multi-echelon effects are not part of the base numerical experiment. The selected shortage penalty is illustrative; in an empirical study it should be estimated from actual lost-margin, expediting, backorder, or customer-retention consequences. The service-level definition must also be selected carefully because cycle service and fill rate measure different aspects of availability [14].

10. CONCLUSION AND FUTURE SCOPE

This study developed a mathematical optimization framework for minimizing relevant inventory costs while improving customer service levels. The model combines ordering, holding, and shortage costs with order quantity and reorder point decisions and imposes an explicit service-level constraint. The numerical experiment demonstrates how systematic enumeration of feasible Q–R policies can identify a cost-efficient policy that satisfies a prescribed service target.

The principal methodological value of the framework is its transparency. Each parameter has an operational interpretation, the feasible region can be inspected, and the resulting policy can be compared with an EOQ-based baseline. Future work should calibrate the framework using actual organizational transaction data, incorporate variable lead times and non-stationary demand, compare probabilistic and robust formulations, and extend the model to multi-item and multi-echelon systems. Further research can also formulate the problem as a multi-objective optimization model that generates a Pareto frontier between cost and customer service rather than selecting a single service target.

REFERENCES

- Harris, F. W. (1913). How many parts to make at once. *Factory, The Magazine of Management*, 10(2), 135–136, 152.
- Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and Production Management in Supply Chains* (4th ed.). CRC Press.
- Hadley, G., & Whitin, T. M. (1961). A family of inventory models. *Management Science*, 7(4), 351–371. <https://doi.org/10.1287/mnsc.7.4.351>
- Hadley, G., & Whitin, T. M. (1962). A family of dynamic inventory models. *Management Science*, 8(4), 458–469. <https://doi.org/10.1287/mnsc.8.4.458>
- Bookbinder, J. H., & Tan, J.-Y. (1988). Strategies for the probabilistic lot-sizing problem with service-level constraints. *Management Science*, 34(9), 1096–1108. <https://doi.org/10.1287/mnsc.34.9.1096>
- Bertsimas, D., & Paschalidis, I. Ch. (2001). Probabilistic service level guarantees in make-to-stock manufacturing systems. *Operations Research*, 49(1), 119–133. <https://doi.org/10.1287/opre.49.1.119.11183>
- Bertsimas, D., & Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations Research*, 54(1), 150–168. <https://doi.org/10.1287/opre.1050.0238>
- Mamani, H., Nassiri, S., & Wagner, M. R. (2017). Closed-form solutions for robust inventory management. *Management Science*, 63(5), 1625–1643. <https://doi.org/10.1287/mnsc.2015.2391>
- Lim, Y. F., & Wang, C. (2017). Inventory management based on target-oriented robust optimization. *Management Science*, 63(12), 4409–4427. <https://doi.org/10.1287/mnsc.2016.2565>
- Inderfurth, K., & Vogelgesang, S. (2013). Concepts for safety stock determination under stochastic demand and different types of random production yield. *European Journal of Operational Research*, 224(2), 293–301. <https://doi.org/10.1016/j.ejor.2012.07.040>
- Snyder, L. V., & Shen, Z.-J. M. (2019). *Fundamentals of Supply Chain Theory* (2nd ed.). Wiley.
- de Kok, T., Grob, C., Laumanns, M., Minner, S., Rambau, J., & Schade, K. (2018). A typology and literature review on stochastic multi-echelon inventory models. *European Journal of Operational Research*, 269(3), 955–983. <https://doi.org/10.1016/j.ejor.2018.02.047>
- Anderson, E. T., Fitzsimons, G. J., & Simester, D. (2006). Measuring and mitigating the costs of stockouts. *Management Science*, 52(11), 1751–1763. <https://doi.org/10.1287/mnsc.1060.0577>
- Jiang, Y., Shi, C., & Shen, S. (2019). Service level constrained inventory systems. *Production and Operations Management*, 28(9), 2365–2389. <https://doi.org/10.1111/>

poms.13060

15. Sadeghi, J., Sadeghi, S., & Niaki, S. T. A. (2024). Multi-objective continuous review inventory policy using MOPSO and TOPSIS methods. *Computers & Operations Research*, 163, 106512. <https://doi.org/10.1016/j.cor.2023.106512>
16. Solyali, O., Cordeau, J.-F., & Laporte, G. (2012). Robust inventory routing under demand uncertainty. *Transportation Science*, 46(3), 327–340. <https://doi.org/10.1287/trsc.1110.0387>
17. Feng, Y., Che, A., & Tian, N. (2024). Robust inventory routing problem under uncertain demand and risk-averse criterion. *Omega*, 127, 103082. <https://doi.org/10.1016/j.omega.2024.103082>
18. Agra, A. (2023). A robust model for the lot-sizing problem with uncertain demands. *Optimization Letters*, 17, 1279–1293. <https://doi.org/10.1007/s11590-023-01994-x>
19. Vernimmen, B., Dullaert, W., Willemé, P., & Witlox, F. (2008). Using the inventory-theoretic framework to determine cost-minimizing supply strategies in a stochastic setting. *International Journal of Production Economics*, 115(1), 248–259. <https://doi.org/10.1016/j.ijpe.2008.06.009>
20. Moon, I., & Choi, S. (1994). The distribution free continuous review inventory system with a service level constraint. *Computers & Industrial Engineering*, 27(1–4), 209–212. [https://doi.org/10.1016/0360-8352\(94\)90272-0](https://doi.org/10.1016/0360-8352(94)90272-0)