FROM DEMAND TO DELIVERY: MATHEMATICAL OPTIMIZATION IN QUICK COMMERCE

(A Quantitative Framework for Inventory, Delivery Time and Operating Cost in India)

Dr. Seema Amit Agarwal, Associate Professor & Vice Principal
Department of Mathematics, Statistics & Computer Systems
Ghanshyamdas Saraf College of Arts & Commerce, Mumbai, Maharashtra, India

Abstract

Quick commerce has changed urban retail by combining local inventory, technology-enabled fulfilment and rapid last-mile delivery. The operational challenge is not simply to deliver quickly, but to balance product availability, delivery capacity and operating cost. This paper develops a mathematical framework for examining these relationships. Inventory requirements are represented as a function of demand, delivery time is modelled as a function of delivery capacity, and a combined cost function is used to demonstrate how differentiation can identify an economically meaningful operating point. A hypothetical numerical illustration shows the method without claiming to describe any particular platform. The present paper focuses on the operational layer of inventory, delivery and cost. The framework is intended as a transparent teaching and research model that can later be estimated using real transaction, inventory and delivery data.

Keywords: Quick Commerce; Inventory Optimization; Delivery Time; Operating Cost; Calculus; Mathematical Modelling; Business Mathematics; Last-Mile Delivery

1. Introduction

Quick commerce, or q-commerce, has developed around the consumer expectation of rapid fulfilment. In India, the operating model relies heavily on dense local fulfilment networks, short delivery radii and coordinated last-mile capacity. Recent industry evidence describes dark stores and last-mile delivery as important parts of the sector's operating structure, while current academic work is also examining delivery-partner selection and order allocation under uncertain demand. This makes quick commerce a useful contemporary setting for applying business mathematics and optimization.

The operational problem is multidimensional. A platform must hold sufficient inventory to satisfy expected demand, but excess stock creates holding and working-capital costs. It must also maintain enough delivery capacity to provide timely service, but excess capacity raises labour and operating expenditure. Consequently, the managerial problem is one of balancing competing costs rather than maximizing any single variable.

Building on mathematical approaches to demand and profitability in digital retail, the present study focuses specifically on the operational dimensions of inventory, delivery capacity and cost."

This paper develops a simplified mathematical framework for that balance. The model is intentionally transparent so that it can be understood by commerce and mathematics students and subsequently extended by researchers using real operational data.

2. Current Business Context in India

Recent market and industry sources describe dark-store-based fulfilment and last-mile delivery as major components of India's quick-commerce logistics system. A 2026 market assessment identifies inventory positioning, labour, delivery cost and store utilization as important operational considerations. A separate 2026 mapping project reports a large and geographically distributed dark-store network across India, illustrating the physical infrastructure behind rapid delivery. These figures provide context and are not used as parameters in the mathematical model.

Academic research is also moving toward more sophisticated operational models. A 2026 article in Expert Systems with Applications proposes an integrated framework for delivery-partner selection and order optimization under uncertain demand. The present paper differs in purpose: it develops a simpler calculus-based framework that can be readily interpreted, taught and empirically extended.

3. Research Problem

The central research problem is: How can a quick-commerce operation balance inventory availability and delivery speed against the costs required to maintain those service levels?

The problem contains two linked trade-offs. First, inventory that is too low increases the risk of unmet demand, while inventory that is too high increases holding cost. Second, additional delivery capacity can reduce delivery time, but each additional unit of capacity has a cost. The study therefore treats the operational problem as a mathematical optimization problem.

4. Objectives of the Study

- To develop a mathematical framework for inventory and delivery decisions in quick commerce.
- To represent the relationship between expected demand and inventory requirements.
- To model delivery time as a function of delivery capacity.
- To examine marginal changes in delivery time and operating cost using calculus.
- To demonstrate a simple optimization model for delivery capacity.
- To develop an inventory-cost formulation that can be extended to real data.
- To provide a research framework suitable for future empirical and computational studies.

5. Research Questions

- How can expected demand be translated into an operational inventory requirement?
- How does delivery capacity affect theoretical delivery time?
- How can holding and shortage costs be represented mathematically?
- How can differentiation identify a potential operating point?
- How can the framework be extended using actual quick-commerce data?

6. Methodology

The study uses a mathematical modelling methodology. It is not an empirical estimation of a particular company's performance. Variables are defined, functional relationships are proposed, derivatives are calculated, and a hypothetical numerical illustration is used to demonstrate the optimization procedure.

The mathematical example is intentionally constructed for illustration. It should not be interpreted as observed data from Blinkit, Zepto, Swiggy Instamart, Flipkart Minutes, BigBasket or any other platform. The framework can be empirically tested later by replacing the illustrative parameters with anonymized operational observations.

7. Variables and Notation

Symbol	Meaning	Unit/Interpretation
D	Expected daily demand	orders or units/day
I	Inventory level	units
N	Delivery capacity	capacity units
T	Average delivery time	minutes
h	Holding cost per unit	₹/unit
p	Cost per unit of unmet demand	₹/unit
F	Fixed delivery-operation cost	₹/day
w	Variable capacity cost	₹/capacity unit
λ	Economic weight assigned to delivery time	₹ per minute-equivalent
C_T	Total operational cost	₹/day

8. Inventory Requirement Model

A simple starting relationship between expected demand and inventory is:

$$I = kD$$

where k is the number of demand-days covered by inventory. For example, if expected demand is 1,000 units per day and k = 1.5:

$$I = 1.5 \times 1{,}000 = 1{,}500 \text{ units}$$

This is a simplified planning relationship. In an actual study, k would be estimated from lead time, demand variability, service-level targets, product characteristics and replenishment frequency.

9. Inventory Holding Cost and Unmet-Demand Cost

If h denotes the holding cost per unit:

$$C_I = hI$$

If demand exceeds available inventory, unmet demand can be represented by:

$$S = \max (D - I, 0) \text{ means } S \text{ is the maximum of } (D - I) \text{ and } 0.$$

Here: S= Shortage or unmet demand; D= Demand; I= Inventory or available stock.

D – I= Difference between demand and inventory

If p is the economic cost assigned to each unit of unmet demand:

$$C_S = p \max (D - I, 0)$$

means Shortage Cost = Penalty per unit × Maximum of (Demand − Inventory) and 0.

Here: C_S= Shortage Cost ; p= Penalty cost per unit of shortage ; D= Demand.

I= Available Inventory ; $\max(D - I, 0)$= Shortage quantity.

These components express the basic trade-off between availability and carrying or shortage-related costs.

10. Delivery Capacity and Delivery Time

A simple decreasing function can represent the relationship between delivery capacity and average delivery time:

$$T(N) = a + b/N$$

where a represents a lower-bound operational component and b captures the workload-related component under the simplified model.

Differentiating gives:

$$T'(N) = -b/N^2$$

For $N > 0$ and $b > 0$, $T'(N)$ is negative. Thus, within the model, increasing delivery capacity is associated with lower delivery time. The absolute marginal effect becomes smaller as N increases, illustrating diminishing marginal improvement.

11. Hypothetical Numerical Illustration

Consider the illustrative delivery-time function:

$$T(N) = 8 + 240/N$$

If N = 10:

$$T(10) = 8 + 240/10 = 32 \text{ minutes}$$

If N = 20:

$$T(20) = 8 + 240/20 = 20 \text{ minutes}$$

The illustration shows the direction of the relationship, not an observed service-time estimate.

12. Marginal Delivery Effect

$$T'(N) = -240/N^2$$

$$T'(10) = -2.4$$

$$T'(20) = -0.6$$

The absolute marginal reduction is smaller at N = 20 than at N = 10. In managerial terms, additional delivery capacity can continue to improve service, but the incremental improvement may diminish.

13. Delivery Cost Function

Suppose the daily cost of delivery capacity is represented by:

$$C_D(N) = F + wN$$

Where:

- $C_D(N)$= Total daily delivery capacity cost.
- F= Fixed daily delivery cost.
- w= Variable cost per delivery unit.
- N= Number of delivery units or delivery capacity.

Explanation:

The total delivery cost consists of two components:

- Fixed Cost (F): Cost that remains constant irrespective of delivery capacity.
- Variable Cost (wN): Cost that increases with the number of delivery units.

Combining speed and cost gives:

$$C(N) = F + wN + \lambda T(N)$$

Substituting the delivery-time function:

$$C(N) = F + \lambda a + wN + \lambda b/N$$

14. Calculus-Based Optimization

$$C'(N) = w - \lambda b/N^2$$

An interior optimum occurs when:

$$C'(N) = 0$$

$$w = \lambda b/N^2$$

$$N^2 = \lambda b/w$$

$$N^* = \sqrt{(\lambda b/w)}$$

The second derivative is:

$$C''(N) = 2\lambda b/N^3$$

For $\lambda > 0$, $b > 0$ and $N > 0$, $C''(N)$ is positive. Therefore, the stationary point is a minimum of the simplified continuous cost function.

15. Numerical Optimization Example

Assume, purely for illustration, that $\lambda = 10$, $b = 240$ and $w = 6$.

$$N^* = \sqrt{(10 \times 240 / 6)} = \sqrt{400} = 20$$

The result indicates an illustrative cost-minimizing capacity of 20 units under the stated assumptions. Because actual capacity is discrete and real operations involve constraints, a practical decision would require integer constraints, staffing availability, demand uncertainty and service-level requirements.

16. Inventory Optimization Extension

A second simplified cost formulation can be written as:

$$C(I) = hI + A/I$$

Differentiating:

$$C'(I) = h - A/I^2$$

Setting the derivative to zero:

$$h = A/I^2$$

$$I^* = \sqrt{(A/h)}$$

This is a mathematical illustration of balancing an increasing holding-cost component against a decreasing cost component. It is not intended to replace established inventory models such as EOQ, safety-stock or stochastic inventory formulations.

17. Integrated Operational Model

Total Cost = Investment Cost + Service Cost + Delivery Cost + Time-related Cost.

Total Cost Function : $C_T(I, N) = C_I(I) + C_S(I) + C_D(N) + \lambda T(N)$

Optimization Problem

The corresponding optimization problem can be expressed as:

$$\boxed{\min_{I,N} C_T(I, N)}$$

Subject to the following operational constraints:

$$I \geq I_{\min}$$

$$N \geq N_{\min}$$

$$T(N) \leq T_{\max}$$

Meaning of Symbols

- I= Inventory level
- N= Delivery capacity or number of deliveries
- $I_{\min}$= Minimum inventory requirement
- $N_{\min}$= Minimum delivery capacity

- $T(N)$= Delivery time function
- $T_{\max}$= Maximum permissible delivery time
- λ= Cost coefficient associated with delivery time

The framework connects demand, inventory, delivery capacity, service time and operating cost without assuming that any single variable should simply be maximized.

18. Conceptual Framework

CUSTOMER DEMAND
↓
INVENTORY REQUIREMENT
↓
AVAILABILITY ↔ HOLDING COST
↓
STOCK-OUT RISK
↓
ORDER FULFILMENT
↓
DELIVERY CAPACITY
↓
DELIVERY TIME
↓
OPERATING COST
↓
MATHEMATICAL OPTIMIZATION
↓
OPERATIONAL DECISION

The framework is conceptual and has not been statistically validated in this study.

19. Discussion

The model highlights three operational principles. First, inventory decisions involve a trade-off between availability and carrying cost. Second, delivery capacity can reduce delivery time, but the model implies diminishing marginal improvement. Third, optimization balances competing objectives rather than maximizing speed or inventory independently.

These principles are consistent with the broader operational structure reported in current industry literature, where dark-store density, inventory positioning, labour, order volume and last-mile efficiency are important elements of quick-commerce economics. The mathematical model converts these qualitative relationships into a form that can be tested with operational observations.

The framework also complements recent academic work that treats quick-commerce fulfilment as a multi-stage decision problem under uncertainty. The present approach is intentionally simpler and more accessible, making it suitable as a foundation for later empirical or computational models.

20. Managerial Implications

- Inventory should be planned in relation to expected demand and replenishment conditions rather than maximized without regard to cost.

- Delivery capacity should be evaluated using marginal service improvement and marginal operating cost.
- Service-speed targets can be incorporated as explicit constraints rather than treated as an unlimited objective.
- Dark-store and fulfilment decisions can be connected to local demand density and delivery capacity.
- Mathematical models can provide a transparent basis for scenario analysis before more complex optimization or machine-learning systems are introduced.

21. Relevance to Commerce and Mathematics Education

Quick commerce provides a contemporary context for teaching business mathematics. Functions can represent demand or delivery relationships; derivatives can explain marginal change; optimization can represent managerial decision-making; and statistics can later be introduced for demand forecasting and model validation.

Mathematical concept	**Illustrative business application**
Function	Demand, cost or delivery-time relationship
Derivative	Marginal cost or marginal service effect
Second derivative	Minimum/maximum analysis
Optimization	Capacity and inventory decisions
Constraints	Service-level and operational limits
Statistics	Forecasting and empirical validation
Operations research	Multi-variable resource allocation

22. Limitations

- The numerical parameters are hypothetical and are not company-reported observations.
- The delivery-time function is deliberately simplified.
- Demand uncertainty is not explicitly modelled.
- Traffic, weather, batching, cancellations and rider availability are not included.
- Product perishability and category-specific inventory characteristics are not incorporated.
- The model does not estimate the actual profitability of any named platform.
- The framework has not been validated against transaction-level or delivery-level data.

23. Scope for Future Research

The framework can be developed into an empirical study by collecting anonymised data on order volume, inventory, delivery distance, fulfilment time, rider availability and operating cost. Future models can incorporate stochastic demand, safety stock, route optimization, multi-product inventory, traffic conditions and service-level constraints.

A future study can use regression or time-series methods to estimate demand and delivery functions before applying optimization. Another extension is multi-objective optimization in which cost, delivery time, stock-out risk and customer service are considered simultaneously.

24. Conclusion

This study develops a calculus-based framework for the operational side of quick commerce. Unlike a demand-pricing-profit study, the present paper focuses on inventory, delivery capacity, service time and operating cost. The mathematical formulations show how derivatives and optimization can be used to examine trade-offs that arise in rapid fulfilment.

The framework should be regarded as a transparent starting point rather than a complete operational model. Its principal value lies in connecting mathematical reasoning with a contemporary business problem and creating a structure that can later be estimated and tested using real data. In this way, the study provides a distinct continuation of research on the mathematics of quick commerce.

References

1. CareEdge Advisory Research. (2025). India's Rapid Delivery Race: Q-Commerce Set to Triple by 2028. CareEdge Advisory Report.
2. Mordor Intelligence. (2026). India Quick Commerce Logistics Market: Industry Analysis, Size and Forecast.
3. Mordor Intelligence. (2026). India Quick Commerce Market: Industry Analysis, Size and Forecast.
4. Singh, N. K., Lahri, V., & Virmani, N. (2026). An integrated framework for delivery partner selection and order optimization in quick commerce under uncertain demand. Expert Systems with Applications, 307, 131078. https://doi.org/10.1016/j.eswa.2025.131078
5. QuickCommerceMap Research Team. (2026). India Quick Commerce Map 2026: 5,625 Dark Stores Mapped.
6. Nexdigm. (2026). India Quick Commerce Market Report.

Appendix A: Summary of Mathematical Relationships

Component	**Relationship**	**Purpose**
Inventory requirement	$I = kD$	Links inventory to expected demand
Holding cost	$C_I = hI$	Represents inventory carrying cost
Unmet demand	$S = \max(D - I, 0)$	Represents stock-out exposure
Stock-out cost	$C_S = p \max(D - I, 0)$	Represents economic impact of unmet demand
Delivery time	$T(N) = a + b/N$	Represents capacity-time relationship

Delivery derivative	$T'(N) = -b/N^2$	Measures marginal service improvement
Delivery objective	$C(N)=F+\lambda a+wN+\lambda b/N$	Balances capacity cost and time
Optimal capacity	$N^* = \sqrt{(\lambda b/w)}$	Illustrative continuous optimum
Inventory extension	$I^* = \sqrt{(A/h)}$	Illustrative inventory-cost balance