

Mathematical Optimization of Product Pricing and Inventory in China Vending Machine Market

Songting Hu

1661 Shengxin North Road, Jiading District, Shanghai, China
sa@cis.ac.cn

Abstract: The rapid expansion of automated retail has made vending machines an increasingly important channel for convenient consumer services in China. However, operators face a joint decision problem involving product pricing, inventory allocation, replenishment, and uncertain demand. This paper develops a mathematical optimization framework for product pricing and inventory management in the Chinese vending machine market. A demand function is constructed to capture the effects of price, location, time, product characteristics, and random demand fluctuations. The model integrates expected sales revenue, purchasing cost, holding cost, shortage cost, and waste or expiration cost into a unified profit objective. A constrained optimization problem is then formulated to determine product-level prices and inventory quantities under capacity, budget, service-level, and operational constraints. The analysis shows that optimal decisions should balance price sensitivity against inventory availability rather than maximize sales volume alone. Products with high demand uncertainty require more flexible safety-stock policies, while products with relatively stable demand can be managed with tighter inventory levels. Location-specific pricing and differentiated inventory policies can improve profitability when consumer demand varies substantially across sites. The paper provides an analytical basis for vending machine operators to coordinate pricing and inventory decisions and identifies directions for incorporating real-time data and machine-learning forecasting into future optimization models.

Keywords: Vending Machines; Product Pricing; Inventory Optimization; Mathematical Optimization; Demand Forecasting; Automated Retail; China Market.

1. Introduction

China's vending-machine industry has expanded alongside mobile payment, automated retail, and demand for convenient consumption. Machines are now common in transportation hubs, universities, office buildings, residential communities, hospitals, and commercial centers. Their small footprint and relatively low labor requirement make them attractive, but profitability remains highly sensitive to price, product assortment, inventory, replenishment frequency, and location-specific demand.

The central operational problem is not simply to maximize sales. A lower price can increase demand but compress contribution margin and raise stockout risk; excessive inventory can protect availability but increase capital tied up in stock and product waste. This paper therefore develops a compact mathematical framework for jointly determining selling prices and inventory quantities under uncertain demand and finite machine capacity [1, 3, 5, 6].

The framework focuses on three practical questions: how demand changes with price and local conditions, how pricing and inventory should be optimized together, and how managers can translate the analytical results into repeatable operating rules [2, 4, 5].

These questions are particularly important because vending machines operate in environments where consumers often make quick, convenience-oriented decisions. Unlike conventional retail stores, the machine provides limited opportunities for customers to compare alternatives, ask staff for recommendations, or return to the same location at a later time. As a result, a small change in price, product availability, or placement can have a measurable effect on transaction probability. From an optimization perspective, this means that operating decisions should be evaluated at the machine and

product level whenever sufficient transaction data are available. A unified model also helps managers identify situations in which a short-term increase in revenue may reduce longer-term profitability through excessive discounting, frequent stockouts, or avoidable waste.

From an operational perspective, the vending machine differs from a conventional retail store because shelf space is extremely limited and the replenishment decision is separated from the customer purchase decision. A store can often move products between shelves, substitute nearby stock, or rely on staff to respond to temporary shortages, whereas a machine has a fixed physical capacity and may remain inaccessible to the operator for several hours or even an entire day. As a result, a seemingly small pricing error can translate into either persistent stockouts or a large amount of unsold inventory. The economic value of an optimization model is therefore not limited to identifying a single price; it lies in coordinating a sequence of interrelated operating decisions.

The Chinese market also contains substantial heterogeneity across machine locations. A machine placed inside a university may exhibit strong demand during class breaks and weaker demand during holidays, while a machine in an office building may experience a more pronounced weekday pattern. Transportation hubs can generate short purchase windows with relatively high willingness to pay, whereas residential communities may favor stable prices and frequently purchased essentials. These differences motivate a location-sensitive formulation rather than a single national pricing rule. By explicitly including location and time effects in the demand function, the model can adapt recommendations to the economic context of each machine.

2. Demand Model and Decision Variables

Let i denote a product, j a vending-machine location, and t a decision period. The main decisions are selling price p_{ij} and initial inventory q_{ij} . Location, time, product characteristics, weather, promotions, and other observable factors are summarized in the baseline demand component α_{ijt} .

$$D_{ijt}(p_{ij}) = \alpha_{ijt} - \beta_i p_{ij} + \varepsilon_{ijt} \quad (1)$$

Here, $\beta_i > 0$ measures price sensitivity and ε_{ijt} is a random demand shock with mean zero and standard deviation σ_{ijt} . A convenient specification for the baseline demand is:

$$\alpha_{ijt} = \exp(\beta_0 + \beta^L L_j + \beta^T T_t + \beta^S S_i + \beta^W W_t + \beta^M M_i) \quad (2)$$

Where L_j represents location characteristics, T_t time effects, S_i product characteristics, W_t external conditions such as weather, and M_i marketing or promotion effects. The exponential form keeps α positive while allowing the same framework to be estimated with transaction data.

Because realized sales cannot exceed the available inventory, unit sales are $\min(D_{ijt}, q_{ij})$. This single relationship connects the pricing problem to the inventory problem: a price change changes demand, and the resulting demand distribution changes the economically optimal stock level [1, 5, 6].

The demand formulation also provides a practical way to distinguish observable variation from uncertainty. Location effects may capture differences in foot traffic and customer composition, while time effects can represent weekday, weekend, holiday, or hour-of-day patterns. Product characteristics may include package size, category, brand strength, and shelf life. Weather and promotional variables are especially relevant for products such as bottled drinks, snacks, and seasonal items. Separating these components is useful because the same product may require a very different pricing and stocking decision at a university, an office building, a transport hub, or a residential community. In addition, the variance parameter σ can be estimated from historical residuals, allowing managers to distinguish high-volume products from products whose demand is difficult to predict.

A further advantage of this formulation is that it separates variables that managers can observe before making a decision from shocks that cannot be predicted perfectly. Location, time, product type, weather, and promotions can be measured or estimated, while ε captures residual uncertainty. This distinction is important for inventory planning because the same expected demand can imply different stocking decisions when uncertainty differs. For example, two machines may each have an expected demand of 30 units, but a machine affected by highly variable foot traffic should normally carry a larger safety stock than a machine with stable and predictable purchases.

The price coefficient β also has a direct managerial interpretation. A larger β indicates that consumers at the corresponding machine are more responsive to price, so a small price increase can generate a relatively large decline in expected demand. A smaller β suggests greater pricing flexibility, although high prices may still be constrained by competition or customer expectations. In empirical applications, β can be estimated from historical transactions by comparing changes in prices with changes in observed sales while controlling for seasonality and other factors. This provides a more defensible basis for pricing decisions than

relying on intuition alone.

3. Joint Pricing–Inventory Optimization Model

Let c_i be unit purchasing cost, h_i holding cost per leftover unit, w_i waste or expiration cost per leftover unit, and s_i shortage cost per unit of unmet demand. The expected one-period operating profit can be written as: [3, 5, 6]

$$\Pi_{ij}(p_{ij}, q_{ij}) = p_{ij} E[\min(D_{ij}, q_{ij})] - c_i q_{ij} - (h_i + w_i) E[(q_{ij} - D_{ij})^+] - s_i E[(D_{ij} - q_{ij})^+] \quad (3)$$

$$\Pi(p, q) = \sum_i \sum_j \Pi_{ij}(p_{ij}, q_{ij}) \quad (4)$$

The objective separates four economic effects: sales revenue, procurement cost, overage cost, and underage cost. This avoids double counting and makes the inventory trade-off explicit. The notation $(x)^+ = \max(x, 0)$ [3, 5, 6].

Operational constraints can be represented by machine capacity, purchasing budget, and feasible price intervals:

$$\sum_i v_i q_{ij} \leq C_j; \quad \sum_i c_i q_{ij} \leq B_j; \quad p_i^{\min} \leq p_{ij} \leq p_i^{\max} \quad (5)$$

Where v_i is the space requirement of product i , C_j is machine capacity, and B_j is the period budget for replenishment. In a multi-period implementation, inventory balance can be added as $q_{ij,t+1} = q_{ij,t} + r_{ij,t} - y_{ij,t}$, with expiry and machine downtime handled as additional constraints.

The constraints also illustrate why maximizing expected revenue is not equivalent to maximizing operating profit. Machine capacity creates an opportunity cost for every unit stocked: allocating space to one product necessarily reduces the space available to another product. A high-margin item may therefore deserve additional capacity even when its absolute sales volume is lower than that of a low-margin item. Similarly, the budget constraint links inventory choices across products and locations, meaning that the optimization problem must determine not only how much to stock, but also where scarce replenishment capital creates the greatest incremental profit.

In a multi-product setting, the model can be extended to account for product substitution. Customers who encounter a stockout of one beverage may purchase another beverage rather than leaving the machine without a purchase. Ignoring this effect can exaggerate shortage costs for some products and underestimate the value of complementary or substitute items. A more detailed formulation could therefore introduce cross-price effects or substitution probabilities, allowing inventory recommendations to reflect the assortment as a portfolio rather than treating each product independently.

4. Analytical Pricing and Inventory Decisions

A closed-form pricing benchmark is obtained by temporarily ignoring inventory scarcity and uncertainty. For expected demand $\alpha - \beta p$, contribution profit is $(p - c)(\alpha - \beta p)$. Differentiating with respect to p gives the interior solution: [4]

$$p^* = (\alpha + \beta c) / (2\beta) \quad (6)$$

The result shows that the optimal benchmark price rises with baseline demand and unit cost and falls with price sensitivity. The solution must still be checked against $p^{\min}$ and $p^{\max}$ and against the effect of inventory constraints.

For inventory, the critical-fractile rule provides a transparent benchmark. Given a price p , the underage cost is

$C_u = p - c + s$ and the overage cost is $C_o = c + h + w$. Therefore: [1, 3, 5, 6]

$$F(q^*) = C_u / (C_u + C_o) = (p - c + s) / (p + h + w + s) \quad (7)$$

If demand is approximately normal with mean $\mu(p)$ and standard deviation σ , the corresponding target inventory is:

$$q^* = \mu(p) + \sigma \Phi^{-1}(F(q^*)) \quad (8)$$

These two expressions make the interdependence clear. A lower price increases expected demand and may also increase the target stock level, while high waste or shortage penalties can materially change the preferred inventory position. In practice, the pricing equation and inventory equation should be solved iteratively or by numerical optimization when capacity, substitutions, promotions, or nonlinear demand are important [1, 4, 5].

An iterative solution is useful because the two decisions influence one another through the demand distribution. Starting from an initial price, the operator can calculate expected demand and a corresponding inventory target. The resulting stock availability can then be used to reassess the expected sales at that price, after which the price can be updated. Repeating this process produces a decision pair that is more consistent with the machine capacity and service objective. This approach is also flexible enough to incorporate operational thresholds, such as minimum margins, maximum acceptable stockout rates, or minimum display quantities for strategically important products. When the model contains multiple products and strong substitution effects, a numerical solver can search across feasible price and inventory combinations while maintaining the required constraints.

The iterative logic can be interpreted as a feedback mechanism. If a lower price increases expected demand substantially, the resulting inventory requirement may consume additional machine capacity and reduce the attractiveness of other products. Conversely, if a product has a high waste cost, the model may recommend a higher price or a lower stocking level to avoid generating demand patterns that create excessive leftovers. This feedback means that pricing cannot be assessed independently from physical availability. The value of an apparently attractive price reduction depends on whether the machine can support the associated increase in demand without creating frequent stockouts or costly replenishment activity.

From a managerial standpoint, the analytical benchmarks are especially useful because they create transparent reference points before a more complicated numerical model is introduced. Managers can first compare the observed price with the benchmark, then evaluate how inventory capacity, uncertainty, and waste alter the recommendation. Such a layered decision process is easier to audit and explain than an opaque algorithm that produces a price without showing the economic mechanisms behind it.

5. Numerical Illustration and Managerial Implications

Consider one product with deterministic demand $D(p) = 120 - 4p$ and unit cost $c = 10$. Ignoring inventory limits, profit is $\Pi(p) = (p - 10)(120 - 4p)$. The first-order condition is $160 - 8p = 0$, giving $p^* = 20$ and demand $D(20) = 40$ units. Profit at this benchmark is 400 monetary units per period.

The example is intentionally simple: it shows that

maximizing unit sales would not produce the same decision as maximizing profit. A price above 20 raises margin but reduces demand too strongly; a price below 20 increases volume but gives up too much margin. Once uncertainty is introduced, the 40-unit demand forecast should be converted into a service-level-based inventory target rather than treated as a guaranteed sales quantity.

For operators, the analytical framework supports three practical rules. First, estimate price sensitivity separately by product and location instead of applying one market-wide elasticity. Second, treat fast-moving, high-margin products differently from slow-moving or short-shelf-life products. Third, re-optimize periodically using transaction data so that price and inventory decisions respond to changes in demand rather than relying on fixed rules [2, 4, 5].

The numerical example can also be interpreted in terms of sensitivity. The demand slope of four units per monetary unit means that a one-unit increase in price reduces deterministic demand by four units. The operator therefore faces a clear trade-off between additional contribution per sale and lost sales volume. In a real machine, this trade-off would be further affected by the probability of stockout, the time until the next replenishment visit, and the remaining shelf life of the product. The simple benchmark should thus be understood as a starting point for a broader operating policy rather than a final recommendation.

An additional managerial implication is that price changes should be evaluated against incremental profit instead of revenue growth alone. A promotion that increases transactions by 20 percent can still reduce profitability if the average contribution margin falls sharply or if additional replenishment trips create significant operating costs. Conversely, a modest price increase may improve profit when demand is relatively inelastic and inventory is frequently constrained. The optimization framework makes these trade-offs explicit and supports performance evaluation at the product-location level.

6. Implementation and Limitations

A feasible implementation is a rolling optimization process: forecast demand for each machine and product, estimate the effect of price on demand, solve the constrained pricing-inventory model, execute the recommended decisions, and update the parameters using newly observed transactions. Historical sales, prices, time, location, promotions, weather, and machine availability can be used as explanatory variables [2, 5, 6].

The rolling process can be implemented on different planning horizons. A daily model can update prices and replenishment quantities for rapidly changing demand, while a weekly or monthly review can be used for assortment and parameter updates. The key principle is to revise decisions when the information set changes materially. Transaction records can be linked with machine identifiers, timestamps, product positions, and observed availability, allowing operators to compare predicted demand with realized sales and identify systematic forecast errors. This feedback loop is important for improving both pricing decisions and inventory accuracy.

The framework is intentionally compact. Real vending systems may include product substitution, promotions, nonlinear demand, replenishment routing, machine downtime, payment failures, competitive responses, and multiple inventory periods. These features can be added through

stochastic programming, simulation, mixed-integer optimization, or machine-learning-based demand forecasts. Empirical validation with transaction-level data from Chinese vending-machine networks remains an important next step [2, 3, 5, 6].

Another limitation is that the present formulation treats the machine and its local market largely in isolation. In reality, nearby vending machines, convenience stores, supermarkets, and food-delivery platforms may influence customer behavior. Competitive price responses can therefore weaken the relationship between a machine price and its observed demand. Product substitution within the same machine creates a similar issue because a stockout of one beverage can shift demand toward another rather than producing a complete loss of the transaction. Future research can extend the model by incorporating cross-price elasticities, spatial competition, dynamic learning, and explicit routing costs. These extensions would make the optimization framework more representative of large-scale vending networks while retaining the same core principle of coordinating demand, price, and inventory.

A practical deployment should also include a governance mechanism for model updates. Pricing recommendations can be reviewed on a fixed schedule, while exceptional events such as severe weather, major promotions, or sudden machine downtime can trigger an earlier re-optimization. A useful system would record the recommended price and quantity, the action actually executed by the operator, and the resulting sales and inventory outcomes. These observations create a continuous feedback loop and make it possible to evaluate whether the model is systematically overestimating demand, underestimating waste, or assigning too much inventory to particular locations.

Another implementation issue concerns data quality. Vending-machine transaction records may contain missing observations, payment failures, temporary machine outages, or abnormal sales spikes caused by events. If these observations are treated as ordinary demand, the estimated price sensitivity and forecast variance can be distorted. Data preprocessing, anomaly detection, and explicit machine-availability indicators are therefore important complements to the optimization model. These practical requirements do not change the mathematical objective, but they strongly influence whether the resulting recommendations are reliable in daily operations.

The framework also assumes that purchasing cost and operating conditions can be represented adequately within the chosen period. In practice, wholesale prices, transportation costs, and supplier terms may change over time. Incorporating time-varying procurement costs and replenishment lead times would allow the model to recognize situations in which a temporary purchasing opportunity justifies additional stock. Likewise, expected demand can change around holidays, public events, and school calendars, suggesting that the forecasting component should be adaptive rather than based solely on historical averages.

7. Conclusion

This paper develops a unified mathematical framework for pricing and inventory decisions in the Chinese vending-machine market. The model links price-sensitive demand, uncertain sales, procurement cost, overage cost, shortage cost,

machine capacity, and operating budget in a single decision structure.

The main implication is straightforward: profitable vending operations require coordination between price and stock rather than independent optimization of sales volume or inventory. The analytical price benchmark and critical-fractile inventory rule provide interpretable starting points, while numerical optimization can incorporate the operational complexity of real machines and locations [1, 4, 5].

Beyond the immediate operating decision, the framework suggests that vending-machine management can be treated as a continuous learning problem. As new transaction observations become available, managers can refine demand parameters, identify persistent location effects, and detect changes in consumer response to price. This creates an opportunity to move from static rules toward adaptive policies that respond to evidence. The value of the proposed framework therefore lies not only in the closed-form benchmark formulas, but also in providing a structured basis for integrating forecasting, optimization, and operational feedback. For Chinese vending-machine operators, such coordination can support more disciplined pricing, better use of limited machine capacity, and a more balanced trade-off between product availability and inventory cost.

The broader contribution of the framework is that it connects two decisions that are often managed separately in practice. Pricing affects not only the revenue generated by each sale but also the volume and uncertainty that the inventory system must absorb. Inventory availability, in turn, affects the effective value of a price recommendation because a high-demand product cannot generate additional sales once the machine is empty. Coordinating these decisions can therefore improve both economic efficiency and operational stability.

Future research can extend the framework toward real-time optimization in which updated transaction data continuously revise demand estimates. Machine-learning models may improve the prediction of nonlinear demand and heterogeneous customer responses, while stochastic or robust optimization can provide more conservative decisions when forecasts are uncertain. Empirical testing with transaction-level data from multiple Chinese cities and machine types would also help determine how strongly the identified relationships vary across locations, products, and operating environments.

References

- [1] Arrow, K. J., Harris, T., & Marchak, J. (1951). Optimal inventory policy. *Econometrica*, 19(3), 250–272. <https://doi.org/10.2307/1906813>
- [2] Chopra, S., & Meindl, P. (2019). *Supply chain management: Strategy, planning, and operation* (7th ed.). Pearson.
- [3] Hadley, G., & Whitin, T. M. (1963). *Analysis of inventory systems*. Prentice-Hall.
- [4] Phillips, R. L. (2005). *Pricing and revenue optimization*. Stanford University Press.
- [5] Silver, E. A., Pyke, D. F., & Thomas, D. J. (2017). *Inventory and production management in supply chains* (4th ed.). CRC Press.
- [6] Zipkin, P. H. (2000). *Foundations of inventory management*. McGraw-Hill.