Resource Allocation with Lumpy Demand: To Speed or Not to Speed?

Track: Operations Planning, Scheduling and Control

Abstract

Given multiple products with unique lumpy demand patterns, this paper explores the determination of both the lot size for each product and the resource allocation among the products, given an investment budget for production rate improvements. Each product's optimal production policy takes on only one of two forms: either continuous production or lot-for-lot production. A heuristic procedure decomposes the problem into a mixed integer program and a nonlinear convex resource allocation problem. The model can be extended to allow the firm to simultaneously alter both the production rates and the incoming demand lot sizes through quantity discounts.

by Bintong Chen and Charles L. Munson
Washington State University

Proceedings of the Twelfth Annual Conference of the Production and Operations Management Society, POM-2001, March 30-April 2, 2001, Orlando, FL.

1. *Introduction*

In this paper we examine the alteration of the production rate, a parameter that is often fixed in inventory models. Munson and Rosenblatt (2001) show that the benefits from coordinated lot sizing among firms in a supply chain are highest when the production rate (fixed in their model) is either low (approaching the demand rate) or high (approaching infinity). This suggests that it may be fruitful to consider the production rate as a variable in inventory models. Production rates can be increased by adding workers, modernizing equipment, adding machines, adding shifts, using overtime, providing training, increasing maintenance, etc. (Bretthauer and Côté, 1996). Researchers have studied changing production rates in various settings (see also Chakravarty and Shtub, 1992; and Jones, Moses, and Zydiak, 1998).

Our research explores the effect of changing production rates on a generalized version of the traditional Economic Production Quantity (EPQ) model that incorporates lumpy demand patterns. We further consider a firm producing multiple products, each with its own lumpy demand pattern. Given an investment budget for productivity improvements, we determine the proper resource allocation among products that are identified as "candidates" for resource allocation. Other papers that have examined the allocation of production resources among products include Bitran and Hax (1981), DeCroix and Arreola-Risa (1998), Evans (1967), Glasserman (1996), and Sox and Muckstadt (1996).

2. *Optimal Production Policy for a Single Product with Lumpy Demand*

Consider the traditional EPQ model, which introduces a finite production rate into the Economic Order Quantity (EOQ) framework. The well-known optimal lot size and associated minimum cost for the EPQ model are respectively given by $\sqrt{2DS/[h(1-D/P)]}$, and $\sqrt{2DSh(1-D/p)}$, where D is the annual demand, S is the setup cost per order, h is the annual holding cost per unit, and P is the annual production rate. Typically the production rate P is predetermined. However, if the company could choose the production rate level, then holding and setup costs would be minimized (equal to 0) when $P = D$. In other words, the optimal production policy would be "continuous production," where the optimal lot size equals ∞, only an initial setup is needed, and inventory never accumulates. Interestingly, unless the firm needs the resources to produce other products, this observation suggests that the company has no incentive (at least from a holding and setup cost perspective) to search for more efficient operating methods to speed up the production rate.

The EPQ and EOQ models assume continuous uniform incoming demand patterns, which is equivalent to the company constantly receiving infinitely small customer order sizes. This assumption is most appropriate for retailers or companies selling to many customers. However, many manufacturers sell to one or very few large customers, implying that their customer demand is "lumpy," i.e. they receive finite customer order sizes intermittently. Next we investigate whether or not the above production rate observation for the EPQ model holds in the lumpy demand case.

We assume one buyer ordering the same quantity Q (perhaps its EOQ) at fixed time intervals. Note that our model is also applicable individually to multiple buyers ordering customer-specific products according to the increasingly common industry practice of "mass customization," where each product needs to be produced separately or orders from different customers are never combined. It can be shown that under this scenario, the optimal lot size for the manufacturer should be an integer multiple n of Q. We assume further that the production rate is a function of investment in technology. For notional convenience, we define the production rate to be $P(x) = DM(x)$, where x is the amount of technology investment used to increase the production rate and $M(x)$ is a multiplier function that is assumed to be concave and strictly increasing with respect to x. In addition, we set $M(0) = 1$, i.e., the current production rate is assumed to equal the customer demand rate. Notice that the above assumptions on $M(x)$ are quite general, and many production functions commonly used in the microeconomics literature satisfy the assumptions. For example, the multiplier function may take the form of $M(x) = 1 + \alpha x^{\beta}$, $\alpha > 0, 0 < \beta \leq 1$. If β=1, $M(x)$ reduces to a linear function of x. The preceding function assumes that an infinite investment leads to an infinitely fast production rate. On the other hand, if the production rate is bounded from above, the multiplier function could take the form of $M(x) = 1 + \alpha(1 - e^{\beta x}), \alpha, \beta > 0$, where the maximum production rate is $(1 + \alpha)D$.

Substituting $DM(x)$ for P and applying Joglekar's (1988) formula, we obtain the optimal holding and setup cost function for the manufacturer when facing lumpy demand of size Q: $c(x,n) = DS/(nQ) + [(n-1) - (n-2)/M(x)]Qh/2$. We are interested in finding the optimal investment $x \geq 0$ and manufacturer's lot size multiplier n to minimize the cost function $c(x,n)$.

Consider the following two cases depending upon the value of n:

- If $n \geq 2$, the cost function is nonincreasing in x. Therefore, an optimal x is 0 and we have $c(0,n) = DS/(nQ) + Qh/2$. Clearly, the optimal n when $n \geq 2$ is ∞ and $c(0,\infty) = Qh/2$.
- If $n = 1$ (lot-for-lot production), we have $c(x,1) = DS/Q + Qh/[2M(x)]$.

Denote $H = Qh/2$, $\rho = 2DS/(hQ^2) = (EOQ/Q)^2$, and $f(x) = \rho + 1/M(x)$. The optimal holding and setup cost as a function of x then reduces to $c(x) = H \cdot \min\{1, f(x)\}$. The preceding analysis implies that the optimal production policy is either continuous production ($n = \infty$) or lot-for-lot production ($n = 1$). Furthermore, investing to increase the production rate is attractive only in the case of lot-for-lot production.

Let $\tau = \lim_{x \to \infty} 1/M(x)$. Since $M(x)$ is strictly increasing, $1/M(x) \geq \tau$ for all $x \geq 0$. Therefore, a necessary condition for the product to receive investment to increase the production rate and perform lot-for-lot production is:

$$\rho + \tau < 1 \quad \text{or} \quad \rho < 1 - \tau. \tag{1}$$

We call a product that meets this condition a "candidate product."

If $\tau = 0$, which happens when infinite investment leads to infinite production rate, then the necessary condition reduces to: $\rho < 1$. In other words, the manufacture should only consider investing in production rate improvement if its EOQ is smaller than the buyer's order size Q.

3. Optimal Production Policies for Multiple Products

3.1 Resource Allocation Among Candidate Products

We now consider the situation where the manufacturer produces multiple products. A total investment budget of B is available to speed up production. The manufacturer needs to decide the amount of investment to allocate to each product in order to reduce total inventory related cost. The lot size of each product then follows directly from the resource allocation as described in Section 2.

Based on the discussion of Section 2, resource allocation should be considered only among the "candidate products." Let I be the set of all candidate products. The implied optimization problem then reduces to the following simplified resource allocation problem (P):

$$\begin{aligned} \text{Minimize} \quad & \sum_{i \in I} c_i(x_i) \\ \text{subject to} \quad & \sum_i x_i \le B \\ & x_i \ge 0 \quad \forall i \in I, \end{aligned}$$

where subscripts i for each product have been added to the notation of Section 2. Problem (P) is a nonconvex, nonsmooth, nonlinear knapsack problem and is difficult to solve in general.

3.2 Heuristic Procedure for Resource Allocation

Instead of attempting to solve (P) optimally by using an enumeration-based procedure, we are interested in designing an efficient heuristic procedure by taking advantage of the structure of the problem. Our heuristic is a decomposition procedure that consists of the following three steps:

1. Solve a linear approximation of problem (P) to determine a subset of products, J, to receive investment allocation. It turns out that the linear approximation can be formulated as a mixed integer program (MIP).
2. Solve a convex nonlinear resource allocation problem to allocate the budget B among products in J.
3. Apply a greedy heuristic to reduce the size of set J by removing one product from J at a time and re-solving the nonlinear resource allocation problem in Step 2. Repeat this process until the total cost no longer improves.

3.2.1 Linear Approximation of (P)

We need the following parameters to introduce the linear approximation of cost function $c_i(x_i)$ for each product i.

- x_i^b is the break-even point of budget allocation that makes production rate improvement worthwhile for product i. It is obtained by solving the equation $1 = f_i(x)$.
- $m_i f_i'(x_i^b)$ is the slope of $H_i f_i$ evaluated at the break-even point.
- E_i is the maximum possible resource allocation beyond x_i^b in the linear approximation. It occurs at the point where the tangent line from x_i^b reaches the lowest possible cost $H_i f_i(\infty) = H_i(\rho_i + \tau_i)$. Simple calculation shows that $E_i = -H_i(1 - \rho_i - \tau_i)/m_i$.

As shown by Figure 1, the linear approximation replaces the nonlinear portion of $c_i(x_i)$ by a tangent line of $f_i(x)$ between x_i^b and $x_i^b + E_i$, and a flat line thereafter.

See Figure 1.

Define $\xi_i, i \in J$, as a binary variable equal to 1 if product i receives resource allocation and 0 otherwise. Define $\delta_i, i \in J$, as a continuous variable representing extra resources allocated to product i beyond the break-even point x_i^b. By definition, the total resource allocation to product i equals $\delta_i + x_i^b$ if $\xi_i > 0$, and equals 0 if $\xi_i = 0$. With $c_i(x)$ replaced by its linear approximation for each i, the resource allocation problem (P) can be reformulated as the following MIP:

$$\begin{aligned} \text{Minimize} \quad & \sum_{i \in I} H_i + m_i \delta_i \\ \text{subject to} \quad & \sum_{i \in I} \delta_i + x_i^b \xi_i \le B \\ & \delta_i \le E_i \xi_i \quad \forall i \in I \\ & \delta_i \ge 0, \quad \xi_i = \{0,1\} \quad \forall i \in I. \end{aligned}$$

In our heuristic procedure, only products with $\xi_i^* = 1$ in the optimal solution of the MIP will be considered for receiving resource allocation. In addition, the optimal objective function value of the MIP provides a lower bound for problem (*P*).

3.2.2. Convex Nonlinear Resource Allocation

Let J be a subset of I such that $J = \{j \in I : \xi_j^* = 1\}$. By restricting resource allocation only among products in set J we may simplify problem (*P*) to a nonlinear resource allocation problem.

Chen and Munson (2001) describe the technical conditions under which a large variety of multiplier functions may be solved. Here we provide results for two special cases that result in closed-form solutions: *linear* and *exponential*.

- If $M_j(x) = 1 + \alpha_j x$ (**linear** form) for all $j \in J$, the resource allocation is:

$$x_j^*(J) = \frac{\sqrt{H_j/\alpha_j}}{\sum_{i\in J}\sqrt{H_i/\alpha_i}}(B + \sum_{i\in J}\frac{1}{\alpha_i}) - \frac{1}{\alpha_j}, \quad j \in J. \tag{2}$$

- If $M_j(x) = 1 + \alpha_j(1 - e^{-\beta_j x})$ (**exponential** form) for all $j \in J$, the resource allocation is:

$$x_j^*(J) = \frac{1}{\beta_j}\left\{\ln\left[\frac{\alpha_j}{(1+\alpha_j)^2}\right] + 2\ln\left(\sqrt{\frac{\beta_j H_j}{4\tilde{\lambda}(J)}} + \sqrt{\frac{\beta_j H_j}{4\tilde{\lambda}(J)} + 1 + \alpha_j}\right)\right\}, \tag{3}$$

where $\lambda^*(J)$ is the unique solution of the following nonlinear equation with respect to λ:

$$\sum_{i\in J}\frac{1}{\beta_i}\left\{\ln\left[\frac{\alpha_i}{(1+\alpha_i)^2}\right] + 2\ln\left(\sqrt{\frac{\beta_i H_i}{4\lambda}} + \sqrt{\frac{\beta_i H_i}{4\lambda} + 1 + \alpha_i}\right)\right\} = B. \tag{4}$$

Notice that the left hand side of equation (4) is monotonically decreasing in λ. Thus many efficient search techniques exist to locate $\lambda^*(J)$. Further note that some solutions to equations (2) and (3) could theoretically be negative. If this happens, the set J cannot be the optimal set of products to receive resource allocation in problem (*P*) and will be eliminated in the refinement step of our heuristic to be described in the next subsection.

3.2.3. Refinements of Resource Allocation

Denote x^*, I^*, and C^* as the optimal resource allocation to each product in I, the set of products that receive the resource allocation in the optimal solution, and the minimum total cost of (*P*), respectively. Denote NC^*(J) as the corresponding objective value of the nonlinear resource allocation problem. By definition, $C^* = NC^*(I^*) + \sum_{i\in I\setminus I^*} H_i$. Since the maximum resource allocation for each product j in the MIP is limited to E_j, I^* is in general (but not always) a smaller set than J and is very likely contained in J. Indeed, our numerical tests in the next section support this observation. We now present a greedy heuristic procedure attempting to reduce the size of set J towards set I^*.

Step (0) Delete all products j with $x^*(J) \le 0$ from set J and re-solve the nonlinear resource allocation problem. Repeat if necessary.

Step (i) Let $k \in J$ be the product such that the following ratio is the smallest:

$$H_j\{1 - f_j[x_j^*(J)]\}/x_j^*(J), \quad j \in J.$$

Step (ii) Let J_ = J\{k}. If the total cost decreases after product k is removed from set J, i.e.,

$$NC^*(J\_) + \sum_{i\in I\setminus J\_} H_i < NC^*(J) + \sum_{i\in I\setminus J} H_i,$$

we set $J = J\_$ and return to Step (i). Otherwise, we stop the refinement and the current set of products J will receive the resource allocation $x^*(J)$.

Notice that in Step (i) the numerator represents the cost increase if product j does not receive resource allocation $x^*(J)$. Furthermore, the numerator will be negative when $x_i^*(J) < x_i^b$, implying that product j did not receive enough allocation to make production rate improvement worthwhile at all. Step (ii) eliminates the product in this case.

4. *Numerical Examples*

To test the performance of the heuristic, we conducted four numerical studies of 50 random samples each. Both linear (Equation (2)) and exponential (Equations (3) and (4)) investment multiplier functions were used for both a 10-product and a 15-product scenario. Optimal solutions were obtained by performing full enumeration over all possible combinations of products in set J containing the products receiving allocation. We chose our study sizes because full enumeration CPU times were about 30 seconds for the 15-product case, and would approximately double for each product added after that. The random data sets were generated in Excel, and the solutions were performed in GAMS/OSL2.

We used the following parameters: H_i = Triangular(50, 100000, 25000), ρ_i = Triangular(.01, .99, .75), B = Uniform(1/10, 1)$\cdot \sum_i (E_i + x_i^b)$ [10-product case], and B = Uniform(1/15, 0.8) $\cdot \sum_i (E_i + x_i^b)$ [15-product case]. For the linear multiplier functions, α_i = Triangular(.01, .00001, .001). For the exponential multiplier functions, α_i = Discrete Uniform(1, 99), and β_i = Triangular(.00001, .001, .0001).

In support of Step 3 of the heuristic (the refinement step), the MIP solution usually contained more products in the allocation set than the optimal solution did. Apparently this occurs because the tangent line reaches the lower bound on cost more quickly than the actual cost function $c_i(x_i)$ does. Furthermore, in most cases (but not all), the products receiving allocation in the optimal solution were included in the MIP solution's allocation set. Thus, the greedy heuristic step of eliminating the product with the least gain per dollar spent usually works well and eliminates the appropriate products. Step 3 has shown to be an important addition to the heuristic.

The heuristic itself performed quite well. Out of all 200 samples, the maximum cost deviation of the heuristic was 2.89%, and the average was 0.0825%. Furthermore, the heuristic provided the optimal solution in 175 samples (87.5% of the time). For comparison purposes, we also calculated the costs of the MIP allocation (the actual values of δ_i^* from the MIP), eliminating steps 2 and 3 from the heuristic. The resulting allocations were generally quite poor, with an average deviation from optimal of 10.1425%. This suggests that while the MIP set (the ξ_i^* values) provides an excellent starting point, the actual MIP allocations are not good, necessitating the solution of the convex nonlinear resource allocation problem (Step 2 of the heuristic).

5. *Extensions*

5.1 Selecting the Best Total Resource Allocation Level B

Following numerous other resource allocation models in the literature, we have assumed that the budget level B is predetermined and will be fully allocated. However, one common drawback of these models is that they usually do not justify the resource allocation, i.e., they do not determine whether the benefit resulting from the resource allocation outweighs the investment itself. Our model can be extended to incorporate this issue in one of the two ways.

First, the user can solve the model for iterative levels of budget B up to a certain capital budget limit. For each budget level, the net present value (NPV) of annual savings (net of B) are calculated over the appropriate time horizon. The budget level yielding the highest NPV is then chosen.

Second, one may alternatively include the resource allocation itself as a separate annualized cost in the objective function. Chen and Munson (2001) explain how the heuristic procedure proposed in this paper can be easily modified to incorporate this extension. The downside of making this change, however, is that any closed-form solutions for x are lost, even if M is a linear function.

5.2 Simultaneously Altering P and Q via Quantity Discounts

For a number of years, researchers have examined the benefits, in the form of lower holding and setup costs for manufacturers, of using quantity discounts to entice customers to change the size of their orders. This technique effectively coordinates the supply chain by finding lot sizes that minimize total system holding and setup costs. In the preceding analysis, we assumed that incoming order sizes Q were fixed. Chen and Munson (2001) show how the model and heuristic in this paper can be extended to incorporate the possibility of changing Q along with the production rate P.

6. Conclusion

In this paper we have examined the EPQ model with lumpy demand to determine under what conditions that production rates should be increased. We have also presented an effective heuristic for allocating limited resources among products that are candidates for productivity improvement.

The answer to the question, “To speed or not to speed,” depends largely on the relative holding and setup costs of the manufacturer and buyer. In general, large incoming order sizes suggest that the manufacturer should use lot-for-lot production and “speed up” as much as possible. On the other hand, small incoming order sizes suggest that the manufacturer should use continuous production and not “speed up” at all, i.e. only produce at a rate (nearly) equal to the demand rate. Moreover, to the extent that the size of incoming orders can be altered, costs will be further reduced as follows. Large orders (when productivity is increased) should be made even larger, and small orders (when productivity is not increased) should be made even smaller. Chen and Munson (2001) shows how the model in this paper can accommodate the simultaneous determination of both production rate and incoming order size.

In a sense, both production policies implied by the model in this paper have a JIT “flavor.” A firm using continuous production produces no faster than necessary. And a firm using lot-for-lot production (hopefully) applies a very high production rate to the process, implying no buildup of inventory. Production is delayed until the last possible moment. Finally, the manufacturer's answer to a *customer* that is ordering JIT (small order quantities), is to *slow down* its production rate (using the continuous production policy). This policy is perhaps counter-intuitive to the idea that a manufacturer should utilize extremely high production rates with JIT customers to be able to respond rapidly to customer orders.

References

Bitran, G.R. and A.C. Hax. “Disaggregation and Resource Allocation Using Convex Knapsack Problems with Bounded Variables”. *Management Science*. Vol. 27, No. 4 (1981). pp. 431-441.

Bretthauer, K.M. and M.J. Côté. “Nonlinear Programming for Multiperiod Capacity Planning in a Manufacturing System”. *European Journal of Operational Research*. Vol. 96 (1996). pp. 167-179.

Chakravarty, A.K. and A. Shtub. “The Effect of Learning on the Operation of Mixed-Model Assembly Lines”. *Production and Operations Management*. Vol. 1, No. 2 (1992). pp. 198-211.

Chen, B. and C.L. Munson. “Resource Allocation in a Lumpy Demand Environment”. Working Paper. Washington State University, 2001.

DeCroix, G.A. and A. Arreola-Risa. “Optimal Production and Inventory Policy for Multiple Products under Resource Constraints”. *Management Science*. Vol. 44, No. 7 (1998). pp. 950-961.

Evans, R. “Inventory Control of a Multiproduct System with a Limited Production Resource”. *Naval Research Logistics Quarterly*. Vol. 14 (1967). pp. 173-184.

Glasserman, P. “Allocating Production Capacity Among Multiple Products”. *Operations Research*. Vol. 44 (1996). pp. 724-734.

Joglekar, P.N. “Comments on ‘A Quantity Discount Pricing Model to Increase Vendor Profits’”. *Management Science*. Vol. 34 (1988). pp. 1391-1398.

Jones, P.C., L.N. Moses, and J.L. Zydiak. “Inventory Investment, Product Cycles, and the Imperfectly Competitive Firm”. *International Journal of Production Economics*. Vol. 54, No. 3 (1998). pp. 267-276.

Munson, C.L. and M.J. Rosenblatt. “Coordinating a Three-Level Supply Chain with Quantity Discounts”. *IIE Transactions*. Forthcoming, 2001.

Sox, C.R. and J.A. Muckstadt. “Multi-Item, Multi-Period Production Planning with Uncertain Demand”. *IIE Transactions*. Vol. 28 (1996). pp. 891-900.

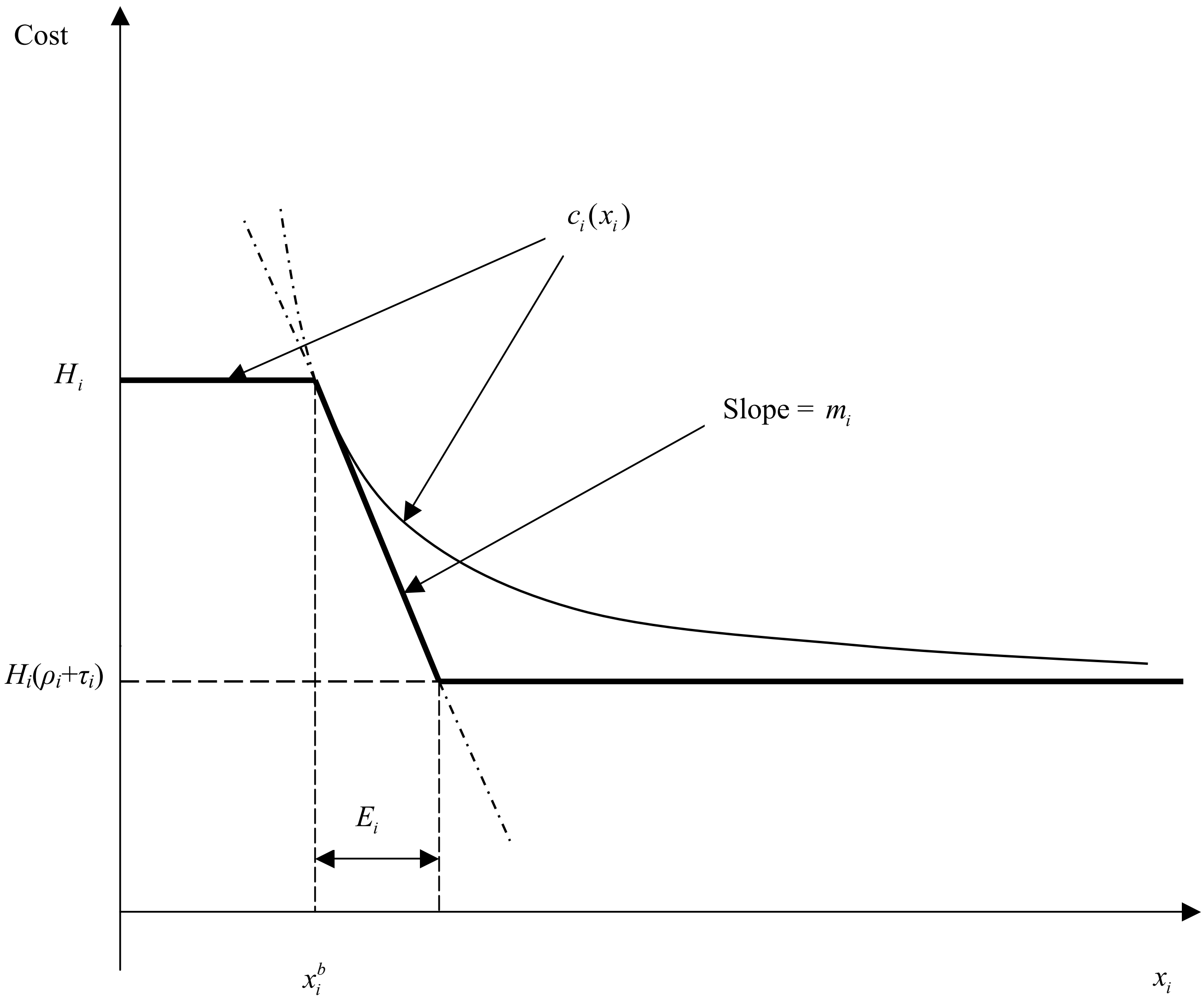

Figure 1. Cost Function $c_i(x_i)$ and Its Linear Approximation