MARGINAL COSTS AND INVESTMENT UNDER UNCERTAIN DEMAND CONDITIONS[1]

by Yong J. Yoon
Public Service Commission of the
District of Columbia

October 1992

New telecommunications technologies such as digital switches and fiber optics allow telephone companies to provide a wide range of new services. However, the provision of these new technologies complicates the regulation of telephony, especially the pricing of services. When a local telephone company provides both basic services and new competitive services, there are incentives for cross-subsidies and predatory pricing[2]. The telephone company may set prices below costs for the new services and try to recover these costs from ratepayers in regulated basic services.

An inevitable question in addressing such issues is the calculation of the marginal cost of services. For some regulated telecommunications companies, long run marginal (incremental) costs (LRIC)[3] are used as a basis for efficient pricing as well as for examining cross subsidies and predatory pricing. Since most services in the telecommunications network require lumpy and recursive investment, the relevant cost for pricing services has often been the long run marginal cost which allows adjustment of all input factors.

In this paper I present a model for deriving long run marginal costs for services requiring lumpy investments. Although the model is developed for investment in digital switches, the analysis can be extended to other lumpy investment such as airports, bridges, highways, and other infrastructure.

Most telephone companies use the capacity cost method as a surrogate for long run marginal cost. The capacity cost method is simple in principle. The marginal unit for the capacity cost is the marginal equipment or the capacity of the marginal equipment.

---

[1] The views expressed in the paper do not necessarily reflect those of the Commission or other Commission Staff.

[2] See Faulhaber (1975) on cross subsidies issue.

[3] We use the terms long run incremental cost (LRIC) and long run marginal cost interchangeably.

The capacity cost of service is the ratio of the outlay on the marginal equipment divided by its capacity; capacity is the maximum unit of service the equipment can provide.

Advocates of the capacity cost method argue that the marginal cost analysis does not apply to telephony because of the complexity of the communications network, and that the capacity cost concept is a good approximation for long run marginal cost. Also, some authors claim that capacity cost is exactly equal to the long run marginal cost under reasonable economic assumptions[4]. Thus, it is worth clarifying the linkage between capacity cost and long run marginal cost.

In essence, capacity cost assumes that services are homogeneous and that demand is perfectly predictable. Thus, the usefulness of the capacity cost concept is limited when services differ in their demand variability or uncertainty. The demand for new services is more uncertain and volatile than the demand for the basic services. Consequently, investment decisions under demand uncertainty would be different from the more certain cases. This observation is relevant for pricing services because long run marginal costs and investment decisions are jointly determined.

Consider a telephone company that provides services on a going concern basis. It faces recursive and lumpy investments because most services in the telecommunications industry require lumpy investments: the large investment of switched network is an example. Then, the relevant cost is the forward looking opportunity cost or the long run marginal cost.

We show that LRIC of a service depends on the demand uncertainty. Usually, the demand for the basic services is stable, while the demand for new competitive services is expected to be uncertain. We show that under this condition, it is more costly to meet the demand for new service than to serve the basic service, plain old telephone service (POTS). Consequently, LRIC of new or competitive services should be higher than that of basic services[5].

---

[4] For instance, see Foster and Bowman (1989). Footnote 6 has a summary of their argument.

[5] As Stigler (1939) pointed out, plant may be planned to provide flexibility of output even if this involves higher costs. For instance, consider the inventory management of a grocery store. Since demand for groceries fluctuates during the week and the seasons, store managers keep inventories. Store managers pay when goods are delivered to the store, and the interest cost is incurred until the good is sold. This interest cost is part of the price consumers pay. Thus, goods with fluctuating demand will incur higher interest costs and price markups than those of goods with predictable demand.

In Section I, Boiteux's peak-load pricing is discussed which introduces the capacity cost method. Section II develops a model for recursive investments for a community where the demand for services grows over time. By allowing all possible adjustments in input factors (plant facilities and operating costs), LRIC reflects the firm's intertemporal decisions. The model introduces the demand uncertainty and the firm's investment decision explicitly. Consequently, investment schedules and long run marginal costs are jointly determined.

Section III derives the long run incremental cost (LRIC) for a reasonable investment strategy under uncertain demand. Conditions are examined under which the capacity cost is a good approximation for LRIC. These conditions indicate that capacity cost can be a poor surrogate for the long run marginal cost (LRIC) for investments that require high fixed costs and face demand uncertainty. Section IV provides concluding remarks.

I. Capacity Costs

This section discusses the capacity cost concept by Marcel Boiteux and others. The main idea is summarized in Boiteux (1960). We start with an example that will show the deceptive simplicity of the "capacity cost." Suppose one unit of equipment (say a digital switch) costs one million dollars and can serve a maximum of 10,000 units of services (say access lines). For simplicity, operating costs are ignored at this level of analysis. The capacity cost of access line service is simply \$100 per service: \$1,000,000÷10,000 = \$100.

This may be a reasonable way of computing the unit cost of service when demand remains constant. However, if demand grows over time, ultimately equipment must be added to meet the growth. If the equipment comes only in large capacity units, then investment will be lumpy and recurring. This is a realistic assumption for investment in modern digital switches, highways and other infrastructure. Unlike in the static case, unused capacity may exist for some of the time, but the capacity cost cannot tell the difference: the capacity cost is still \$100 because the outlay and capacity of the equipment have not changed. The question is then, what is the cost of serving additional demand when there is unused capacity?

The answer will depend on how the additional demand is viewed. One response to this question has been that the marginal cost is zero, and the additional service should be free (assuming zero operating costs). This would be the case if the additional demand is temporary and the duration of the additional service is short enough that the placement schedule of investment remains undisturbed. But if the additional demand is permanent, the placement schedule will be affected, and the demand would cause additional costs to the firm. The long run marginal costs under this condition of demand are presented in Section II.

Boiteux maintains that marginal costs differ according to whether they are planned to produce the extra unit "once-only" or to permanently raise by one unit the "flow" of output. His example of "extra passenger" is instructive. A train is about to leave, and there is one empty seat. A passenger arrives. The cost of carrying this extra passenger is zero (again assuming zero operation cost). But the same argument is valid for all the empty seats that there may be in the train. Then, the optimum rate as is understood by the marginal theory is zero.

To correct the argument of "extra passenger," Boiteux treats the railway car, not the seat, as the marginal unit, while implicitly assuming that an optimum size of investment is chosen. Thus follows Boiteux's conclusion that the service must be priced at marginal development cost or the capacity cost.

II. Model for Recursive Lumpy Investment

A dynamic model for lumpy investment is formulated in this section, and the long run marginal costs are derived under uncertainty in the following Section III. In developing models for pricing public utilities, Turvey (1969) emphasized that both cost and output have time dimensions, and that both may be subject to uncertainty. Thus, for a cost analysis to be useful in decision-making, it has to be dynamic (intertemporal) rather than static.

Turvey assumes that all equipment is fully utilized. This may be a reasonable assumption in many cases, but would be inappropriate for investment in telecommunications infrastructure which we wish to analyze, because demand for service is growing over time, and new equipment has to be added to meet the growing demand. If the available capacity unit is large relative to the service unit, then the investment is lumpy and new capacity is underutilized. Future demand will be met by recursive investment. Given a long-term forecast for demand, a cost minimizing capacity expansion path is generated, and the total cost, the present value of the whole future costs, can be calculated.

Long run marginal cost (LRIC) is calculated by considering two scenarios. Consider an alternative path of demand growth, and calculate the total cost of alternative capacity expansion. The difference between the two costs is the incremental cost in meeting new demand path.

To be more formal, consider two scenarios: the "baseline" service is growing over time (scenario 0), and the "alternative" (scenario 1) is a higher level of service growing at the same rate as the baseline (see Figure 1). We assume that the demand for services grows at a rate g (units of service per period). Then, the demand at time t is

$$(1) \qquad D_t = D_0 + gt$$

where, for convenience, we assume the initial demand D_0 is zero. To meet the steady growth of demand, there corresponds a schedule of investment separated by regular time intervals, which we call the "placement interval" (see Figure 1). (Note that we described the demand without any consideration of price change. This can be justified by the principle of stable ratemaking.)

Now the supply side is described. Let "I" denote the outlay on the equipment; the capacity of the equipment is Q units of services. For convenience, assume the equipment lasts forever, without depreciating. This assumption does not affect the result. We also assume that there is no technological change.

Then, the "placement interval" N is given by

$$N = Q \div g$$

and the time schedule of adding equipment is

$$t_0 = 0,\ t_1 = N,\ t_2 = 2N,\ \ldots$$

The total cost (TC) of the baseline scenario is the present value of investment outlays, given the discount rate r,

$$TC(0) = \sum_{i=0}^{\infty} I \div (1+r)^{iN} = I * \{1 + 1 \div [(1+r)^{N} - 1]\}.$$

Suppose there is a new arrival of demand, which causes a permanent demand increment by m units for each period. For convenience, the incremental demand as a fraction f of the capacity is

$$m = fQ,\ 0 < f < 1.$$

The demand flow of the alternative schedule is

$$D'_t = fQ + gt.$$

The alternative scenario requires placement intervals move forward by fN periods (see Figure 1),

$$(2) \quad 0,\ t_1 = N - fN,\ t_2 = 2N - fN,\ \ldots$$

The total cost of the alternative investment schedule is

$$TC(1) = I*\{1 + 1\div(1+r)^{N-fN} + 1\div(1+r)^{2N-fN} + ... \}$$

$$= I*\{1 + (1+r)^{fN} \div[(1+r)^{N} - 1]\}.$$

Then the incremental cost (IC) is the difference between the two total costs, the present values of the two investment schedules:

$$(3) \qquad IC = TC(1) - TC(0)$$

$$= I*[(1+ r)^{fN} - 1]\div[(1+r)^{N} - 1].$$

The present values usually apply to cash flows expressed in nominal dollar terms. For non-monetary quantities, time discounting is a dubious practice unless the price is fixed over time. Assuming stable ratemaking, service price remains a constant. This assumption justifies present values of demand, though the assumption is not part of the marginal cost theory.

Based on this qualification, the present value of the demand increment is

$$(4) \qquad dD = fQ*\{1 + 1\div(1+r) + 1\div(1+r)^{2} + ... \} = (fQ)(1+r)/r.$$

III. LONG RUN INCREMENTAL COST

The long run incremental cost (LRIC) is simply equation (3) divided by equation (4),

$$(5) \qquad LRIC = IC \div dD = \frac{rI \quad \{(1+r)^{fN}-1\}}{(1+r)fQ\{(1+r)^{N}-1\}} \quad .$$

In computing LRIC it was assumed that new demand arrives at the time of new investment (at time 0). However, new demand may arrive any moment while new equipment will have unused capacity. In this case, the LRIC of services may change over the life of the equipment. When the equipment is first installed, capacity is relatively abundant and the shadow price of capacity is low; whereas, as the unused capacity is filled, the shadow price is higher[6]. LRIC is higher right before, than right after the installation of the equipment.

Thus, it is desirable to express LRIC as a function of the arrival

---

[6] Note that the shadow price is not zero even if there is excess capacity. In a static model the shadow price is zero if there is excess capacity.

time and size of new demand. Equation (5) provides the incremental cost right after the installation, and this is denoted by LRIC(0) if necessary to avoid confusion.

Suppose a new demand (in the same amount fQ) arrives at time t and stays for good. The present value of the new demand will be

$$dD_t' = (1+r)^{-t} dD_t$$

$$= (1+r)^{-t} fQ(1+r)/r$$

where dD_t comes from equation (4). If a new demand of size fQ arrives while there is available capacity, the new demand will be met by existing capacity. This is the case if the arrival time t is in the interval $0 \le t \le N-fN$. The investment plan is according to equation (2), and the LRIC is the incremental cost in equation (3) divided by the incremental demand dD_t' above,

$$\text{(6)} \qquad LRIC(t) = IC \div dD_t'$$

$$= (1+r)^t LRIC(0) \qquad \text{for } 0 < t < N-fN.$$

The long run marginal cost in equation (6) is greater than the LRIC in equation (5) because of the accumulated interest cost until t when new demand arrives.

It should be noted that formula (6) holds for new demand arriving between time periods 0 and N-fN, when there is sufficient capacity to serve new demand. What if new demand arrives when there is not sufficient capacity (i.e., the available capacity is less than fQ)?

If the company knows the exact timing of new demand, or if new demand can be served by an immediate addition of new capacity, an optimal strategy would be to install new equipment immediately upon new demand[7]. This assumption is not realistic, however, because

---

[7]. This is the assumption made by Foster and Bowman (1989). Under this assumption Foster and Bowman show that LRIC is equal to capacity cost. It is not difficult to provide an intuitive explanation. In Foster and Bowman, new demand is met either by the existing capacity or by an immediate addition. In either case, new demand does not incur interest costs that is the opportunity cost of any additional capacity above status quo. Consequently, LRIC is exactly equal to the capacity cost.

The investment plan of Foster and Bowman scenario depends on the arrival time. If the new demand arrives within the time interval $0 < t < N-fN$, then the investment plan is the same as equation (2) in Section (II), adding new capacity in infrastructure like a digital switch requires adequate lead time and uncertain demand requires capacity installed in advance.

To analyze investment plans under uncertain demand, we assume random arrivals of new demand. Under such conditions it is well known in management science that the firm maintains inventories or spare capacity to meet demand fluctuations[8]. Stigler (1939) also observes that plant may be planned to provide flexibility of output even if this causes additional costs.

The appropriate investment strategy under demand uncertainty is to install a new capacity (equipment) when the remaining capacity is fQ, the size of additional demand, whether or not new demand has actually arrived. This strategy reflects the firm's "going concern" nature and optimizing behavior under uncertainty as well.

(A.1) 0, N-fN, 2N-fN, ...

If the new demand arrives after N-fN, then the remaining capacity is not sufficient to meet the new demand; the size of new demand is fQ. In this case, Foster and Bowman assumes that new equipment will be installed immediately. Suppose new demand arrives at t after time N-fN. Then, the investment schedule will have to change accordingly,

(A.2) 0, t, 2N - fN, 3N - fN, ...

Following the same procedure used in the paper, the total cost and the incremental cost for Foster-Bowman case can be obtained. Then calculate the ALRIC by integrating over the interval (0, N0) and obtain the Foster-Bowman result

$$\text{(A.5)} \qquad ALRIC = \frac{rI}{(1+r)Q} .$$

[8] See Scarf (1960) and Yoon (1985) for application of (s, S) inventory models in economics. Consider a retailer who faces economies of scale in making orders. The (s, S) policy is characterized by two critical numbers, the low case s and the upper case S. If the inventory is above the level s, no orders are made; when inventory falls below s, then orders are made to bring the inventory level up to S. This feature of (s, S) models is useful in analyzing recursive lumpy investment. An investment in equipment involves a fixed cost and the variable cost that depend on the capacity. The (s, S) optimal policy suggests an optimal size S for the equipment and a deployment decision of installing a new capacity if the remaining capacity is s. This model provides a tractable method of calculating LRIC under uncertain demand conditions more general than the case treated in this paper.

If the firm notices there is insufficient capacity to meet possible new arrival, to avoid the cost of adding capacity on short notice or losing customers (or failing to serve customers in case of going concern), the firm installs new capacity in expectation of the need.

For new demand arriving during the period N-fN and N, LRIC can be obtained by comparing the two total costs. The placement schedule is exactly the same as that for the "alternative" case in equation (2), and the same formula (4) applies for demand (see figure 2).

Assuming that the arrival of new demand is uniformly distributed over the N periods, the time average of long run incremental cost (ALRIC) can be calculated. The ALRIC can be considered as a relevant cost for stable rate making:

$$(7) \qquad ALRIC = \frac{1}{N}\int_0^N LRIC(t)\,dt$$

$$= \frac{rI*[(1+r)^{fN}-1]*[(1+r)^N-1]}{N*\ln(1+r)\ \ (1+r)*fQ*[(1+r)^N-1]}$$

$$= \frac{rI}{(1+r)*Q}\ \frac{[(1+r)^{fN}-1]}{fN*\ln(1+r)}\ .$$

We note that ALRIC is a product of annualized capacity cost, rI/(1+r)Q, and the correction factor. For an infinitely small demand increment, ALRIC becomes the annualized capacity cost. By taking the limit as the fraction f goes to zero, we obtain

$$(8) \qquad \lim_{f \to 0} ALRIC = [rI \div (1+r)] \div Q$$

$$= \text{Annualized cost of investment} \div \text{capacity}$$

$$= \text{Annual capacity cost,}$$

where rI÷(1+r) is the annualized cost of the investment. The implication of the result in (8) is that the cost of serving a small demand increment is approximately the capacity cost.

IV. Critique of the Capacity Cost Method

A common practice in telecommunications costing has been the

capacity cost method. The standard argument in support of the capacity cost method has been that capacity cost is a good surrogate for long run incremental cost (LRIC) because it is theoretically sound and administratively simple. Perhaps this would be the case if capacity could be added in small size.
We note that technological conditions in telecommunications determines not only the kinds of new services to be provided, but also the market structure and the adequacy of economic cost concepts.

In Section III it was shown that LRIC is not equal to capacity cost unless demand is known with certainty. New competitive services are expected to face uncertain demand conditions than are plain old telephone service (POTS). The uncertain demand condition is expressed by the random arrival and the size of new demand, which reflects the variation of demand over the trend.

Conditions under which capacity cost is a good surrogate for LRIC can be obtained by examining equation (7), which expresses ALRIC as a product of capacity cost and the correction factor. The conditions demonstrate limitations of the capacity cost concept as costing method for investment in infrastructure. The capacity cost method will be a reliable device for calculating LRIC only under the following limited conditions:

(1) the placement period (N) is short;
(2) the size of demand increment (f) is negligible;
(3) there is no technological change.

Condition (1) implies that the capacity of equipment is small relative to the demand growth. Investment in durable infrastructure usually fails to meet this condition. Condition (2) indicates that the demand is stable and predictable, which would be the case for plain old telephone services (POTS), but not for new competitive services. Condition (3) matters, but is not discussed here.

Since the long run marginal cost from equation (7) is a product of capacity cost and the correction factor, the discrepancy between marginal cost and capacity cost can be measured by the quantity (correction factor - 1). This quantity tells the accuracy of capacity cost as a surrogate for long run marginal cost.

The accuracy or the error margin of capacity cost is reported in Table 1. Therein, capacity Q is unity (1), and the discount rate is 10% (r = 0.1). The growth rate in the baseline is denoted by g, and the replacement interval is N = Q/g. The table shows that the error margin increases with the summary statistic fN. The term fN can be rewritten as fN = fQ/g: the size of the demand variation relative to the rate (g) of general demand growth. This is not surprising because fN is a measure of demand uncertainty in the model, which can be interpreted as the portion of demand that is not captured by the trend.

V. Conclusion

The paper has developed a model for recursive and lumpy investment in infrastructure, and derived the long run marginal cost (ALRIC) of service under uncertain demand. The linkage to capacity cost is also obtained by expressing ALRIC as a product of capacity cost and the correction factor.

Two kinds of uncertainty are considered in this paper. The size of new demand arrival is a measure of demand uncertainty. Another kind of uncertainty is the timing of arrival. It is assumed the size of arrival is known, but the arrival time is random. Thus, a measure of uncertainty in this paper is the variability of the service away from the trend, which is the baseline.

Under conditions of demand uncertainty, which we believe are important aspects of new telecommunications services, the firm will use an investment strategy that adjusts to anticipated uncertainty. The firm would meet the demand as a going concern; or the firm may try to maximize expected profit by providing some flexibility in the form of inventories or excess capacity. In either case, the strategy is to add new capacity when the firm notices the remaining capacity is not sufficient to meet expected demand.

From the proposed model, the long run incremental cost of service is calculated. The main result is that capacity cost is a good surrogate for LRIC only for stable and predictable services. For a service with uncertain demand, capacity cost can be a poor substitute and needs adjusting. For instance, as is demonstrated in Table 1, if the equipment lasts 20 years (an approximate lifespan of a switch), and the demand fluctuation of a new service is expected to be 10% of the equipment's capacity, then the LRIC is about 11% higher than the capacity cost.

Table 1.a
Q = 1, r = 0.1, N = Q/g.

g \ f	0.05	0.1	0.2	0.4
5 % (N=20 yrs)	1 % (fN=1)	11 % (fN=2)	22 % (fN=4)	50% (fN=8)
10 % (N=10 yrs)	1 % (fN=0.5)	1 % (fN=1)	11 % (fN=2)	22 % (fN=4)
20 % (N=5 yrs)	1 % (fN=0.25)	1 % (fN=0.5)	1 % (fN=1)	11 % (fN=2)
50 % (N=2 yrs)	1 % (fN=0.1)	1 % (fN=0.2)	1 % (fN=0.4)	1 % (fN=0.8)

Table 1.b

fN	Correction
0.1	1 %
0.2	1 %
0.4	1 %
0.8	1 %
1	1 %
2	11 %
4	22 %
8	50 %

References

Boiteux, M. "Peak-Load Pricing," Journal of Business (April 1960): 157-79; also in J. Nelson, Marginal Cost Pricing in Practice (Prentice-Hall, 1964): 59-89.

Ronald Coase, The Firm, the Market, and the Law (The University of Chicago Press, 1988).

G. Faulhaber, "Cross-subsidization: Pricing in Public Enterprises," American Economic Review (1975): 966-977.

Kenneth Flamm, "Technological Advance and Costs: Computers versus Communications," in Changing the Rules (The Brookings Institution, Washington, D.C. 1989).

Richard W. Foster and Robert M. Bowman, "An Incremental Cash Flow Approach to Long Run Cost Analysis," in Telecommunications Costing in a Dynamic Environment (Proceedings of the Bellcore-Bell Canada Conference on in San Diego, California, April 1989).

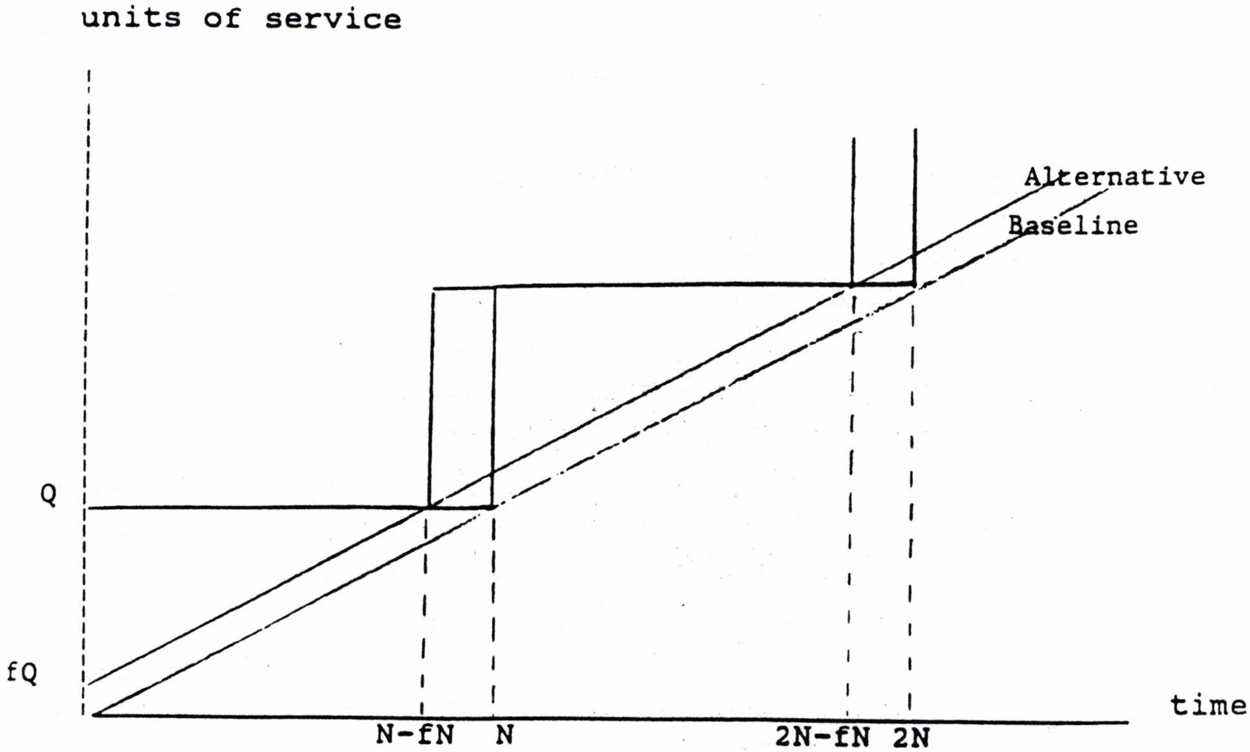
Herbert Scarf, "The Optimality of (s,S) Policies for the Dynamic Inventory Problem," Proceedings of the First Stanford Symposium on Mathematical Methods in the Social Sciences (Stanford University Press, 1960).

George Stigler, "Production and Distribution in the Short Run," Journal of Political Economy (vol. 47, 1939): 304-27.

Ralph Turvey, "Marginal Cost," Economic Journal (June 1969): 282-299.

Y. J. Yoon, "An Aggregate Inventory Equilibrium: (s, S) Policy and Search," mimeo (1985).

Figure 1



Note. N is the placement interval.

Figure 2 The model in this paper

units of service

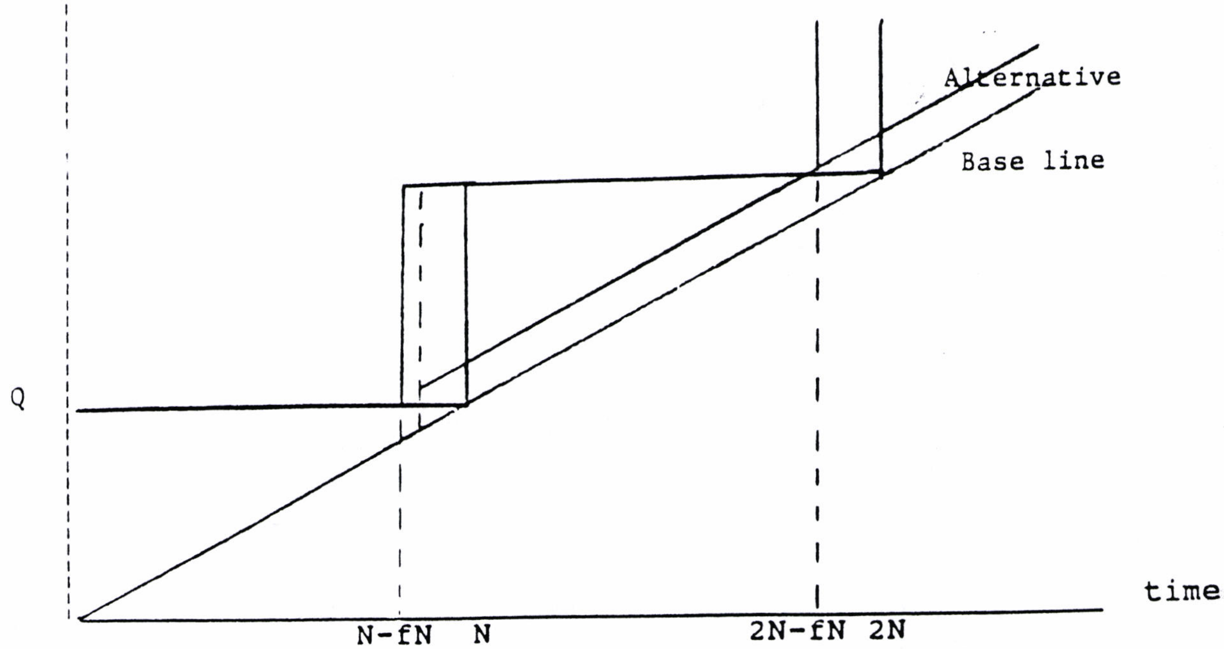


Figure 3. Foster and Bowman model

units of service