Dynamic Peak-Load Pricing*

Oz Shy†

University of Haifa and Stockholm School of Economics

Revised, February 19, 2001

Abstract

This paper develops a dynamic theory of peak-load pricing and service provision. Unlike the literature on peak-load pricing which assumes that the demand for service in each period is predetermined, here, consumers are utility maximizers and are free to postpone buying the service to a different season. In turn, service providers can manipulate the price system to shuffle consumers between seasons as to maximize profit or social welfare.

Results show the conditions under which the classical peak-load pricing algorithm may need to be modified since a firm or a social planner will find less beneficial to delay the service when the time discount factor is taken into account. The paper also shows that peak-off-peak price- or quantity-reversals cannot be realized in an environment where consumers are allowed to postpone buying the service.

Keywords: Dynamic Peak-load pricing, Price discrimination, Seasonal pricing

JEL Classification Numbers: D4, L2, L5, L9

(Draft = peak37.tex 2001/02/19 10:25)

*I thank Ted Bergstrom, Tore Ellingsen, Paul Kleindorfer, Peter Morgan, and Juha Tarkka, as well as seminar participants at the Helsinki School of Economics, Institute for International Economic Studies at Stockholm University, University of Copenhagen, Lund University Center for Operations Research & Econometrics of Université Catholique de Louvain, Stockholm School of Economics, and Tel Aviv and Haifa Universities for most helpful comments.

†Department of Economics, University of Haifa, 31905 Haifa, Israel. E-mail: `ozshy@econ.haifa.ac.il` Homepage: `http://econ.haifa.ac.il/∼ozshy/ozshy.html`

1. Introduction

Services constitute what economists call *nonstorable goods.* Electricity, telephone, transportation, banking and most other services are used (consumed) at the time of purchase. This nonstorability characteristic of services leads to a congestion of service systems when the demand for the service is unevenly distributed among different seasons. The demand for telephone services is the highest during week days and is lower during the nights, weekends, and holidays. The demand for air travel is high during the summer for most places, whereas the demand for transportation to ski resorts is greater during the winter. Electricity use follows a daily cycle, as need for using appliances and lighting pass through a daily routine, and it also follows a yearly cycle due to climatic change. Thus, the demand follows periodic cycles.

The classic peak-load pricing theory, formalized in Boiteux (1960), Steiner (1957), and Williamson (1966) relies on the assumptions that the demand for service in one season is *independent* of the demand in other seasons, see Crew and Kleindorfer (1979, 1986), Brown and Sibley (1986), Sherman (1989), and Crew, Fernando, and Kleindorfer (1995) for literature surveys. The drawback of this theory is that it abstracts from a more general behaviour in which at least some consumers may choose to shift their demand from one season to another in response to a lower price during their "less desirable" season.

It should be pointed out that, in the literature, there are several formulations that introduce substitution. Crew and Kleindorfer (1986, Sec. 3.4), following Dansby (1975), allow for nonzero cross derivatives of a season's demand function with respect to output sold in different seasons. Bergstrom and MacKie-Mason (1991) and Shy (1996, Ch. 13) allow for substitution between seasons in utility-based models. However, these formulations are *static!*

In this paper we do not assume that the demand for service in each period is independent of other seasons. More precisely, we construct a dynamic consumers-overlapping-generations model in which utility-maximizing consumers enter the market in one season and have three options regarding the purchase of the service: (a) buy the service at the time of entry, (b) postpone buying the service to a subsequent period (e.g., next season), and (c) not buying at all. In this framework, the demand for service in each period is endogenously constructed by the relative seasonal pricing of this service. Therefore, by using the price mechanism, service providers can reduce the load in a period when a large number of consumers enter the market. Thus, the present framework allows for the service to be differentiated according to the time of consumption and allows for service substitution between seasons.

A second aspect that is missing from the conventional theory is a dynamic treatment.[1] A peak-load problem is *by definition* a dynamic problem. Postponing service reduces the value of the service to consumers and in addition delays the payment received by the service providers. Thus, a peak-load analysis must model what role the rate of time preference (interest rate) plays in consumers' and producers' decisions.

The main issues analyzed in the present framework are:

1. How do optimal peak-load prices in the dynamic setup compare with the optimal prices prescribed by the "classic" static setup?

2. How do optimal prices vary with the time discount factor (hence the interest rate)?

3. Can a price or a quantity reversal be realized? That is, is there a case where a social planner sets a high price during the period where less consumers enter the market? Can this outcome result in a situation that less service is provided during the season where more consumers enter the market?

The paper is organized as follows. Section 2 describes the dynamic model where consumers can postpone buying the service to a later season and define the welfare criterion. Section 3 characterizes optimal peak-load pricing. Section 4 concludes with discussions of the results.

[1] Crew and Kleindorfer (1979, Ch. 7), following Pressman (1979), formulate a continuous-time capital investment model; they assume predetermined instantaneous demand functions with no possibility of delayed consumption.

2. A Dynamic Model

This section develops the basic model of consumers who can postpone buying the service when they face lower prices in a subsequent period.

Consider an infinite horizon economy. Time is discrete and is denoted by t, $t = 0, 1, 2, 3, \ldots$. Time periods are divided into *seasons*. The even time periods indexed by $t = 0, 2, 4, 6, \ldots$ are called *Summer season*, whereas the odd time periods $t = 1, 3, 5, \ldots$ are called *Winter season*. For every even period t, we define a *cycle* at $T = t$ as the two periods (period t Summer plus period $t + 1$ Winter).

Note that there is no particular reason for assuming that the seasons are called Summer and Winter (hence, a cycle refers to one year). I use these terms for convenience only. The present model can be applied to day and night (phone or electricity services), in which case the cycle is one day. These types of 'seasons' can be distinguished by applying different rates of time preference to the particular problems to be analyzed.

2.1 Consumers

Each consumer lives for two periods. In each even (Summer) period, n^S consumers enter the market; in each odd (Winter) period, n^W consumers enter the market. Each consumer purchases the service at most once, either in the first or the second period of his life, and exits the market forever. With no loss of generality assume that $n^S \geq n^W \geq 0$ meaning that the number of consumers entering during the Winter season is no greater than the number of consumers entering during the Summer season.

In each period t, the entering consumers are indexed by δ on the interval $[0, 1]$ according to an increasing utility towards delaying in obtaining the service. Thus, a consumer indexed by $\delta = 0$ will never delay buying the service whereas a consumer indexed by $\delta = 1$ will delay buying the service if the next season's price is lower.

Let $U_\tau(\delta)$ denote the utility function of a consumer indexed by δ and who enters the market in some period τ and may buy the service either in period τ or period $\tau + 1$. Let ρ

($0 < \rho < 1$) denote the time discount factor. Define

$$U_\tau \stackrel{\text{def}}{=} \begin{cases} v - p_\tau & \text{if he buys the service in period } \tau \text{ (now)} \\ \rho(\delta v - p_{\tau+1}) & \text{if he buys the service in period } \tau+1 \text{ (delayed one period)} \\ v_0 & \text{does not buy any service.} \end{cases} \tag{1}$$

Thus, the parameter δ measures the importance consumer δ attaches to having the service (task) performed 'now' as compared to next season. The parameter v measures each consumer's basic valuation for the service, and the parameter v_0, $0 < v_0 < v$, measures consumers' reservation utility which is the utility derived from obtaining a similar service elsewhere or not getting the service at all.

The parameters ρ and δ in a consumer's utility function (1) must be distinguished. This distinction is very important for our analysis. The parameter ρ is simply a time discount factor and therefore multiplies the entire net utility when a service and the payment are postponed by one period. In contrast, the consumer-specific δ parameter measures the urgency of the task to be performed with this service. Low δ implies that the task must be performed right away (an important business trip, or making an urgent phone call). Notice that the parameter δ does not multiply the price as it affects utility only.

Let $\hat{\delta}_\tau$ denote the period τ entering consumer for whom all period τ entering consumers indexed by $\delta < \hat{\delta}_\tau$ purchase the service with no delay, and all consumers indexed by $\delta > \hat{\delta}_\tau$ postpone the service to the next period (season). From (1), $\hat{\delta}_\tau$ is determined by $v - p_\tau = \rho(\hat{\delta}_\tau v - p_{\tau+1})$, or

$$\hat{\delta}_\tau = \begin{cases} 1 & \text{if } \; p_{\tau+1} \geq \dfrac{p_\tau - (1-\rho)v}{\rho} \\[2ex] \dfrac{v + \rho p_{\tau+1} - p_\tau}{\rho v} & \text{if } \; p_{\tau+1} < \dfrac{p_\tau - (1-\rho)v}{\rho}. \end{cases} \tag{2}$$

Figure 1 illustrates how generation τ entering consumers are divided between consuming 'now' and 'next period.'

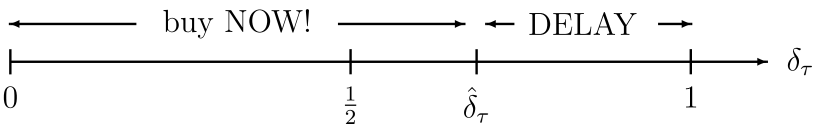

Figure 1: Generation τ consumers' choice of service. All consumers indexed on $[0, \hat{\delta}_\tau]$ buy 'now,' whereas all consumers indexed on $(\delta_\tau, 1]$ delay consumption.

2.2 Production of services

The cost of providing service is divided into two types: Capacity cost and operating expenses. Capacity cost refers to the investment in infrastructure needed for providing a certain amount of service in a given time period. For example, in the airline industry capacity often refers to the number of passenger-seats available to the airline in a given period of time. In the telecommunication industry, capacity refers to the number of transmissions the network can handle in a given time period. In the hotel business, capacity refers to the number of beds. For a restaurant, capacity may mean the number of tables, and so on.

The difference between capacity cost and operating expenses lies in the assumption or fact that firms must commit to their level of capacity *before* they can provide service. Moreover, it is further assumed that the investment in capacity is either irreversible, or that capacity is insufficiently liquid so that a firm cannot sell or lease it during periods of low demand. Clearly, this assumption is too strong. For example, airline companies can temporarily lease aircraft at periods in which they find it profitable to provide extended service. Phone companies can channel phone calls via the lines of other phone companies. In fact, small phone companies in the US rely on the infrastructure belonging to the big three telephone companies. We disregard these observations and assume that for our purposes, firms must commit to pay for the loans they take in order to invest in capacity.

We denote by K the capacity level chosen by a firm before it starts providing the service, and by k the per-unit-capacity payment it makes *each* period (season) whether or not this capacity is utilized.[2] Such payments include interest payments on a loan taken to finance

[2]Here, I deviate from the (static) literature that generally assumes that k is the per-unit capacity cost for *all seasons/periods* of service.

the investment in capacity. We denote by c the operating expenses of one unit of service. For example, in the airline industry the operating expenses is the cost of handling each passenger and the cost of inflight services, such as food. In the hotel business, this cost is of maintaining an occupied room, such as cleaning, electricity, and water.

2.3 Welfare criterion

Since the number of consumers entering the market in each even period n^S is constant, and since the number of consumers entering in each odd period n^W is also constant, we restrict the analysis to stationary outcomes meaning that all cycles yield the same allocations. Thus, we denote the 'summer' price by $p^S = p_t$ for all t even; and the 'winter' price by $p^W = p_t$ for all t odd.

Proposition 1

If in each Summer some of the Summer-entering consumers delay buying the service to the Winter, then no Winter-entering consumer delays buying the service to the Summer; and vice versa. Formally, for every t, if $\hat{\delta}_t < 1$ then $\hat{\delta}_{t+1} = 1$.

Proof. Either $p^S \geq p^W$ or $p^S < p^W$. The utility function (1) implies that if $p^S \geq p^W$, none of the Winter-entering consumers postpones buying the service since $v - p^W \geq \delta v - p^S$ for every $\delta \in [0, 1]$. Similarly, if $p^W \geq p^S$, none of the Summer-entering consumers postpones buying the service since $v - p^S \geq \delta v - p^W$ for every $\delta \in [0, 1]$. □

Proposition 1 implies that a social planner always sets $p^S \geq p^W$. Otherwise, if $p^W > p^S$, consumers entering when a small number of consumers enter the market will delay consumption to a period when more consumers enter the market, and this allocation is clearly dominated. Altogether, $\hat{\delta}_t = \hat{\delta}^W = 0$ for all odd t, and $\hat{\delta}_t = \hat{\delta}^S \geq 0$ for all even t.

We need to establish a welfare criterion. Naturally, the social planner would maximize a discounted sum of generations' utilities, $\sum_\tau \rho^\tau U_\tau$, where U_τ is defined in (1). Due to stationarity, it is more convenient to 'slice' $\sum_\tau \rho^\tau U_\tau$ into cycles instead of generations. Therefore, we define the *per cycle (per term) utility* U_T to be the discounted sum of utilities of those period $t = T$ Summer-entering consumers who are not delayed plus the discounted utilities of Summer and all Winter consumers buying service in period $t = T + 1$. Hence,

$$\begin{aligned} U_T(\hat{\delta}^S;\rho) &\stackrel{\text{def}}{=} \hat{\delta}^S n^S v + \rho n^S v \int_{\hat{\delta}^S}^{1} \delta \, d\delta + \rho n^W v \\ &= \rho v \left[n^W + \frac{n^S}{2} \right] + n^S v \left[\hat{\delta}^S - \frac{\rho (\hat{\delta}^S)^2}{2} \right]. \end{aligned} \tag{3}$$

The per-cycle utility is drawn in Figure 2 below as a function of $\hat{\delta}^S$ (the number of summer-entering consumers who are not delayed). The slope of the per-cycle utility is given by

$$\frac{\partial U_T}{\partial \hat{\delta}^S} = n^S v \left[1 - \rho \hat{\delta}^S \right] > 0, \quad \text{since } \hat{\delta}^S < \frac{1}{\rho}. \tag{4}$$

The total cost per term (cycle) is given by

$$TC_T(\hat{\delta}^S;\rho) = \begin{cases} [(1+\rho)k + c] n^S \hat{\delta}^S + \rho c [(1-\hat{\delta}^S) n^S + n^W] & \text{if } \hat{\delta}^S \geq \bar{\delta} \\ n^S \hat{\delta}^S c + [(1+\rho)k + \rho c][(1-\hat{\delta}^S) n^S + n^W] & \text{if } \hat{\delta}^S < \bar{\delta}, \end{cases} \tag{5}$$

where $\bar{\delta} = 1/2 + n^W/(2n^S)$ is the fraction where total cost is minimized (happens to be also where capacity is minimized). The first row in (5) refers to an allocation where more consumers are served during summers than in winters. In this case, capacity is determined by the summer-service level only, hence capacity costs are $(1+\rho)kn^S\hat{\delta}^S$.

The second row in (5) refers to an allocation where more consumers are served during winters than in summers. In this case, capacity is determined by the winter-service level only, hence capacity costs are $(1+\rho)k[(1-\hat{\delta}^S)n^S + n^W]$. All other costs in (5) are operating costs (which equal c times the service level in a season).

The slope of the per-cycle cost as a function of $\hat{\delta}^S_T$ is

$$\frac{\partial TC_T}{\partial \hat{\delta}^S} = \begin{cases} n^S[(1-\rho)c + (1+\rho)k] & \text{if } \hat{\delta}^S \geq \bar{\delta} \\ n^S[(1-\rho)c - (1+\rho)k] & \text{if } \hat{\delta}^S < \bar{\delta}. \end{cases} \tag{6}$$

The per-cycle total cost function is drawn in Figure 2, again as a function of $\hat{\delta}^S$.

The second part of (6) shows that if the per-unit operating expense is large relative to the per-unit capacity cost (i.e., $k < (1-\rho)c/(1+\rho)$), the total cost function is upward sloping even for small $\hat{\delta}^S$ where the capacity investment is decreasing. This happens because an

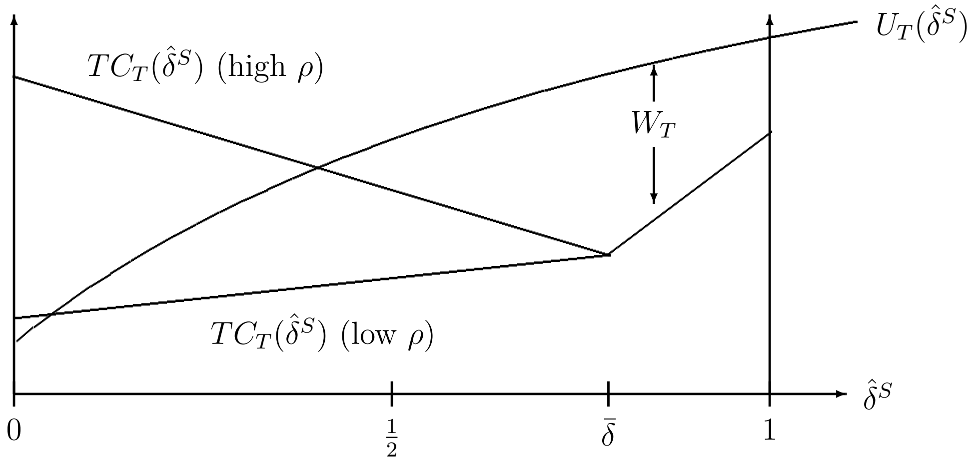

Figure 2: Utility per cycle and total cost per cycle as functions of $\hat{\delta}^S$. *Note:* for $\hat{\delta}^S < \hat{\delta}$, the total cost function is upward sloping for 'high' values of c (or 'low' values of ρ), and downward sloping otherwise.

increase in $\hat{\delta}^S$ reduces the number of consumers delaying service from Summer to Winter; hence decreasing the saving associated with delayed payment of operating expenses. Thus, for large values of c, this increase in operating expense dominates the saving associated with reduced capacity at this range. Clearly, this case is unlikely to hold for ρ sufficiently close to 1.

Finally, we assume that regulator chooses $\hat{\delta}^S$ to

$$\max_{\hat{\delta}^S} W_T(\hat{\delta}^S) \stackrel{\text{def}}{=} U_T(\hat{\delta}^S) - TC_T(\hat{\delta}^S), \tag{7}$$

where $U_T(\hat{\delta}^S)$ is defined in (3), and $TC_T(\hat{\delta}^S)$ is defined in (5). In the next section, we show how the regulator can support the optimal $\hat{\delta}^S$ by a price mechanism.

3. Analysis

3.1 Socially-optimal allocation of service

Figure 2 raises the question whether it is possible to have a *peak reversal* situation where social welfare is maximized when more than half of the total number of Summer and Winter entering consumers buy the service during the Winter. That is, can social optimum be consistent with $\hat{\delta}^S < \bar{\delta}$?

Proposition 2

There is no peak reversal. That is, the number of consumers served during Summers is never below the number of consumers served during Winters. Formally, $\hat{\delta}^S > \bar{\delta}$.

Proof. Suppose that $\hat{\delta}^S < \bar{\delta}$. To rule out a reversal, we need to show that

$$\left.\frac{\partial U_T}{\partial \hat{\delta}^S}\right|_{\hat{\delta}^S<\bar{\delta}} = n^S v\left(1-\rho\hat{\delta}^S\right) > n^S[(1-\rho)c-(1+\rho)k] = \left.\frac{\partial TC_T}{\partial \hat{\delta}^S}\right|_{\hat{\delta}^S<\bar{\delta}},$$

which immediately follows from the assumption that $v > c+2k$ □

Comparing the slopes for the range where $\hat{\delta}^S \geq \bar{\delta}$ yields the socially-optimal $\hat{\delta}^S$.

$$\hat{\delta}^S = \begin{cases} 1 & \text{if } \rho < \dfrac{v-k-c}{v+k-c} \\ \dfrac{v-(1+\rho)k-(1-\rho)c}{\rho v} & \text{if } \dfrac{v-k-c}{v+k-c} < \rho < \dfrac{v-k-c}{\bar{\delta}v+k-c} \\ \bar{\delta} = \dfrac{2n^S+n^W}{2n^S} & \text{if } \rho > \dfrac{v-k-c}{\bar{\delta}v+k-c}. \end{cases} \tag{8}$$

Equation (8) shows that when ρ is sufficiently small social optimum requires no delay in service to all consumers ($\hat{\delta}^S = 1$). Hence, the introduction of a discount factor reduces the need to maintain different seasonal prices. To see this, for the interior solution we have that

$$\frac{\partial \hat{\delta}^S_T}{\partial \rho} = \frac{c+k-v}{\rho^2 v} < 0.$$

Hence,

Proposition 3

An increase in the rate of time preference increases the socially-optimal number of consumers whose service is delayed to the Winter.

Thus, as consumers discount the future less, more Summer-entering consumers will receive delayed service.

3.2 Pricing

Finally, the social planner must utilize an optimal seasonal pricing structure to support the optimal allocation of service $\hat{\delta}^S$. To support the (interior) socially optimal $\hat{\delta}^S$ given in (8), the utility function (3) implies that $v - p^S = \rho(\hat{\delta}^S v - p^W)$, hence for the interior part

$$p^S = (1+\rho)k + (1-\rho)c + \rho p^W. \tag{9}$$

Now, if we apply the classic optimal peak-load pricing mechanism (e.g., Steiner 1957), ignoring the possibility of 'peak reversal,' a social planner in a static planner would set $p^W = c$ and $p^S = 2k + c$. That is, the summer consumers bear the entire capacity costs associated with their cycle.

Suppose that the social planner sets $p^W = c$ just like in the classic algorithm. Then, (9) implies that $p^S = (1 + \rho)k + c < 2k + c$, that is lower than the level prescribed by the peak-load pricing classic algorithm. Therefore,

Proposition 4

The classic algorithm "over charges" the Summer (peak) season consumers.

Thus, the classic algorithm prescribes seasonal prices differences that exceed the socially-optimal price difference, thereby inducing "too many" Summer-entering consumers to delay buying the service to the Winter. Notice that this difference is eliminated when $\rho \to 1$.[3]

4. Conclusion

The contribution of the present paper to the understanding of the peak-load pricing problem can be summarized as follows:

1. Market conditions determine which season is "high" and which season is "low." Previous literature assumed that high and low seasons are exogenously predetermined.

2. Consumers are allowed to substitute between the seasons whenever they find that the price differential effect dominates the loss of utility associated with delaying consumption.

3. The paper analyzes the peak-load pricing problem as a dynamic investment problem rather than as a static one.

[3]Crew and Kleindorfer (1979, Ch. 7) do not obtain this dependency of optimal peak-load prices upon the rate of time preference. The reason for this difference is that their demand functions are instantaneous and do not allow for substitution among the seasons.

4. The model starts out with utility functions (rather than demand functions) therefore allows us to draw exact welfare implications.

The paper introduces two independent modifications to the classical model: First, I allow consumers to delay buying the service to a less desirable season (role played by the consumer-specific δ). Secondly, I introduce a time discount factor into the decision making process (the parameter ρ). As it turns out, the first modification does not alter the basic peak-load pricing algorithm derived under the predetermined demand theory. The paper shows that only the second modification implies that the basic optimal peak-load algorithm should be modified to take into account the time-discount factor. Table 1 compares the consequences of applying the classic peak-load pricing algorithm to a dynamic environment. Table 1 shows

Var.	Classic algorithm	Compared	Dynamic algorithm
p^S	$2k + c$	$>$	$(1+\rho)k + c$
p^W	c	$=$	c
K	$n^S \dfrac{v - 2k - (1-\rho)c}{\rho v}$	$<$	$n^S \dfrac{v - (1+\rho)k - (1-\rho)c}{\rho v}$

Table 1: The consequences of applying the classic algorithm in a dynamic environment. p^S is an optimal 'Summer' price; p^W is an optimal 'Winter' price; K is capacity investment

that when the classic pricing algorithm is implemented, the economy is distorted by having a larger than optimal price gap between the seasons, thereby inducing more than the optimal number of consumers to delay service. Hence, the economy will underinvest in capacity.

Finally, the reader may wonder whether the assumption that all consumers prefer obtaining the service 'now' over 'later' (rather than all preferring, say, Summer over Winter) is restrictive. I would like to argue that it is not since in this model the relative seasonal demand is controlled by the number of entering consumers, n^S and n^W. For example, $n^S > 0$ and $n^W = 0$ should be interpreted as having all consumers prefer the Summer service over Winter service. In contrast, $n^S = 0$ and $n^W > 0$ should be interpreted as having all consumers prefer the Winter service over Summer service. Thus, there is no need to assume that consumers are oriented towards a specific season since this is already captured by consumers' choice in which season to enter.

References

Bergstrom, T., and J. MacKie-Mason, 1991, "Some Simple Analytics of Peak-load Pricing," *Rand Journal of Economics,* 22: 241–249.

Boiteux, M., 1960, "Peak-Load Pricing," *Journal of Business,* 33: 257–179.

Brown, S., and Sibley, D., 1986, *Public Utility Pricing,* Cambridge, England: Cambridge University Press.

Crew, M., and P. Kleindorfer, 1979, *Public Utility Economics,* New York: St. Martin's Press.

Crew, M., and P. Kleindorfer, 1986, *The Economics of Public Utility Regulation,* Cambridge Mass.: The MIT Press.

Crew, M., C. Fernando, and P. Kleindorfer, 1995, "The Theory of Peak-Load Pricing: A Survey," *Journal of Regulatory Economics,* 8: 215–248.

Dansby, R., 1975, "Peak Load Pricing with Time Varying Demands," Bell Labs, Unpublished.

Pressman, I., 1970, "A Mathematical Formulation of the Peak-Load Pricing," *Bell Journal of Economics,* 1 (Autumn): 304–326.

Sherman, R., 1989, *The Regulation of Monopoly,* Cambridge: Cambridge University Press.

Shy, O., 1996, *Industrial Organization: Theory and Applications,* Cambridge, Mass.: The MIT Press.

Steiner, P., 1957, "Peak-loads and Efficient Pricing," *Quarterly Journal of Economics,* 585–610.

Williamson, O., 1966, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraints," *American Economic Review,* 56: 810–827.