Multi-period Congestion Pricing Models and Efficient Tolls in Urban Road Systems

LOUIE NAN LIU *

College of Management, Zhejiang University

Abstract

This paper reviews recent advances in multi-period congestion pricing models in urban road systems. Mathematical formulations of various congestion pricing problems for two time periods (peak and off-peak) and for a simple road network are presented. A procedure is provided for conducting a simulation study of the peak and off-peak congestion pricing models to examine congestion tolls and their effects on traffic allocations and social welfares. Major findings from the analysis results are summarized. Some extensions of the existing models and future research topics are also discussed.

1 Introduction

The economic theory of congestion pricing has long been recognized by economists since Pigou (1920) and Knight (1924). The use of peak-period pricing to improve the use of scarce or time-sensitive resources is a well-established economic principle. The full cost of a trip on a congested road includes not just a traveler’s own time and vehicle operating costs, but also the costs that traveler imposes on all other travelers by adding to the level of congestion. A congestion price can thus be viewed as a user charge that is based on the difference between the cost perceived by the user when entering the traffic stream and the cost actually imposed on all users as a result of the additional delay caused by the user’s entry and movement through the traffic stream.

Congestion pricing, as Downs (1992) indicates, is one of the demand-side strategies that focus on behavior. The principal objective of congestion pricing is to alleviate congestion by implementing surcharges for the use of selected congested facilities during peak traffic periods. By shifting some trips to off-peak periods, to routes away from congested facilities, or to higher-occupancy vehicles, or by discouraging some trips altogether, congestion pricing schemes would result in savings in time and operating costs for both private and commercial vehicles, improvements in air quality, reductions in energy consumption and improvements in transit productivity. Congestion pricing also promises to generate large amounts of new revenues which could be used to provide improved transportation alternatives or for other purposes.

* College of Management, Zhejiang University, 20 Yugu Road, Hangzhou 310027, CHINA. E-mail: nliu@mail.hz.zj.cn
Moreover, recent advances in electronic toll collection (ETC) technologies have made congestion pricing technologically feasible. Unlike traditional toll plazas, toll rates on electronic lanes can be adjusted instantaneously. Electronic toll lanes can handle much higher traffic flows because vehicles need not stop to make payment. Among various ETC applications, it should be mentioned that California State Route 91 (SR-91) was the first urban expressway in the United States that has been subject to time-of-day congestion tolls since December 1995. McDonald et al. (1999) provided a detailed description of the SR-91 case.

Although congestion pricing has been proposed as an effective strategy for tackling traffic congestion, many issues associated with implementation of congestion pricing, as discussed in Downs (1992), remain unsolved. One of the issues is efficiency. The problems of technical complexity, traffic spillover, and the tradition of free access to local streets make the implementation of congestion pricing inherently difficult. Under these technical and/or political constraints, it may not be possible for an electronic toll-charging system to cover all the roads and streets in a metropolitan area; hence, there will be portions of the urban road network that cannot be subjected to efficient tolls.

To study congestion pricing there is a need to develop models allowing for travelers’ choices of departure time, route and mode, as well as whether to travel at all. More specifically, these models should consider a road network and multiple daily time periods (called multi-period models hereafter) so as to analyze the effectiveness of toll policies on travelers’ departure time and route choices. The main advantage of the multi-period model is that it can handle both spatial and temporal effects of congestion pricing. For example, the two-period model (a peak period and an off-peak period) is able to deal with question such as how much congestion can be relieved by shifting peak period traffic to off-peak periods. On the other hand, the two-period model also creates another policy instrument, the toll in the off-peak period.

The above issues involves two different fields: economic theory of peak-load pricing, and second-best pricing of transportation systems. The literature on peak-load pricing is fairly rich but most studies are related to public utilities, for example, telephone utilities. Among others, Mohring (1970) investigated the peak-load problem with increasing returns and pricing constraints. Sherman (1989) provided a detailed overview of peak-load pricing theory. Sherman considered second-best pricing, which is defined as the most efficient prices that are possible given any specific constraint that must be satisfied, such as a budget constraint requiring total revenue to equal total cost. This definition is useful in formulating a second-best congestion pricing problem.

In particular, Pressman (1970) makes a distinct contribution to the formulation of peak-load pricing problems. In this paper, Pressman uses more realistic period demand functions – dependent demand functions, that is, the demand for a good or service depends on all the goods and/or services demanded in the multi-period problem involved. For example, for a regulated telephone utility that sells a single service in two distinct periods known as the peak and off-peak periods, each demand function would depend on both the peak and off-peak demand. With the dependent demand functions, Pressman generalized the notion of consumers’ surplus by using the line-integral calculus. Although the paper is intended for the peak-load pricing problems in public utilities, the ideas also apply to the formulation of the multi-period congestion pricing problem.

On the other hand, the theory of second-best pricing considers the question of efficient prices given some constraint that prevents the use of a complete set of first-best marginal
cost prices. Second-best pricing of roads has been studied by many researchers, including the original work by Levy-Lambert (1968) and Marchand (1968); recent studies by Berstein and El Sanhouri (1994) and Braid (1996) on the use of dynamic bottleneck model to deal with the pricing of transportation facility with an unpriced substitute; work by McDonald (1995) and Verhoef et al. (1996) using a two-link static equilibrium model to study the second-best one-route congestion pricing in the presence of an untolled alternative. Verhoef (2002) extended the single-period second-best congestion pricing model to a general road network. Multi-period congestion pricing model was initially presented in Small (1992) for a single link/route network. Liu (1995) and Liu and McDonald (1998; 1999) developed two-period congestion pricing simulation models to examine second-best tolls for a two-route network. Liu and Boyce (2002) examined the problem of optimal congestion pricing for a general transportation network with multiple time periods.

The purpose of this paper is to review recent advances in multi-period congestion pricing models, particularly those in Liu (1995), Liu and McDonald (1998; 1999), and Liu and Boyce (2002). Following this introduction, Section 2 describes the road system considered in our study, including the network, travel cost and demand characteristics. Section 3 presents mathematical formulations of various congestion pricing problems for two time periods and for a simple road network. Section 4 discusses the solution procedure – a simulation study of the peak and off-peak congestion pricing scheme, and examines congestion tolls and their effects on traffic allocations and social welfares. Section 5 presents some generalizations and extensions of the existing models, and future research topics.

2 Characteristics of the urban road system

2.1 Network representation

Consider an urban road network consisting of two routes connecting an origin (for example, home) and a destination (for example, workplace) over multiple distinct time periods during the day. To be more specific, the study investigates the problem of the morning commute, and considers two time periods – peak and off-peak periods, to describe traveler’s departure time choice. The capacity for each route is assumed to be fixed and to be the same for two periods. Each traveler can travel either early during the pre-peak period, or during the peak period. Without loss of generality, assume the two periods are of equal length and normalized to length one.

The two-route network has been used by other researchers (for example, Marchard, 1968; Ben-Akiva et al., 1986; Arnott et al., 1990; Verhoef et al., 1996; Liu and McDonald, 1999; De Palma and Lindsey, 2000). The main reason for choosing such a system is that the network is simple enough to conduct an economic analysis of traveler’s route choice behavior. The routes considered here have no specific engineering configuration and are assumed to be perfect substitutes. During each period, trip from the origin to the destination can be made on either route.
2.2 Travel costs

Because of the focus of the study is on congestion, only short run costs related to congestion will be considered. The **average cost**, the cost borne by an individual traveler, includes vehicle operating costs and travel time costs which are caused by congestion. For each time period, average costs on each route are assumed to be a monotonically increasing function of the traffic volumes on the route during that period. The **marginal cost** of an additional unit of traffic volume, consists of the average cost and the additional cost that a traveler imposes on all other travelers (on all links and in all time periods) by adding to the level of congestion. The **total cost** consists of the costs borne by all travelers (for each route and for each time period) in the system.

In the system, one route is a toll route; due to technical and/or political constraints, the other route must remain untolled. Tolls are temporally differentiated, that is, peak toll vs. pre-peak toll. Thus, the full travel cost on the toll route for each period is the average cost plus the congestion toll, while the full travel cost on the free route for each period is the same as the average cost.

2.3 Travel demand and formulation of social benefits

Now consider the aggregate hourly travel demand rate between the origin and destination in different time periods over the course of a day. This demand is called **period demand** and is assumed to be dependent, that is, the demand in one period is a function of the prices of traveling (or trip prices) in both peak and pre-peak periods. To simplify the specification of the demand functions, the income effect is assumed to be negligible.

A simple specification for the period demand functions would be a linear relationship between the period demand and the prices in both periods. The relationship assumes that the peak (pre-peak) demand is negatively related to the peak (pre-peak) price of traveling, and positively related to the pre-peak (peak) price of traveling. In the study, assume the inverse period demand function exists, that is, the trip price for one period, can be expressed as a function of traffic volumes in both peak and pre-peak periods.

The purpose of introducing the dependency of period demands is to study the shifting peak problem, that is, the diversion of the peak period trips to the pre-peak period by considering the response of the peak period demand to the pre-peak trip price, and vice versa. However, the dependency may generate two difficulties with the formulation of gross social benefits which here can be given as a line integral of the two inverse period demand functions. The first difficulty relates to the definition of the line integral because it depends on the particular path on which the integral is calculated and is thus not unique. The second one relates to the differentiability of the line integral which is a part of the object function to be optimized.

As discussed in several studies, including Liu and McDonald (1998, 1999), Nagurney (1999), Small (1992), and Williams (1976), both difficulties can be resolved by assuming a symmetry or integrability condition on the inverse demand functions, which requires that the first cross-period derivatives of the inverse peak (pre-peak) demand functions with respect to pre-peak (peak) traffic volumes are equal. This condition says that the effect on the peak period trip price resulting from a change in the pre-peak traffic volume is the same as the effect on the pre-peak trip price resulting from a change in the peak traffic volume.

The above condition is one of the fundamental properties associated with the compensated (or Hicksian, that is, cost-minimizing demand function) demand function.
Since this study assumes the demand function has an income effect of zero, the demand function can be regarded as the Hicksian demand function which satisfies the integrability condition. Under the condition, the line integral is uniquely defined and is independent of the path chosen; and the first derivative of the line integral with respect to the traffic volumes for one period is equal to the inverse period demand function, that is, the trip price for that period. Pressman (1970) provided a complete discussion of this integrability condition.

3 Model formulation of multi-period congestion pricing problems

3.1 Formulation of a second-best congestion pricing problem

Given the theoretical framework in the previous section, the second-best congestion pricing problem (Model SB) can thus be formulated as a constrained optimization program. The objective function is the maximization of the social welfare or net benefits, which is defined as the total benefits (B) minus the total costs (C):

\[ \max (\text{total benefit} - \text{total cost}) \]

The total benefit is defined as a line integral of the inverse peak and pre-peak demand functions. The total cost is the total travel costs (traffic volumes × average travel cost) on each route (toll route and free route) and for each period (peak and pre-peak).

Four constraints are associated with the optimized program: (1) equilibrium trip price condition in the peak period. This is the constraint on the pricing of the free route in the peak period. In the peak period, the equilibrium price of a trip on either route is equal to the average cost on the free route; the equilibrium price of the trip on the toll route is the average cost plus the congestion toll in the peak period. (2) equilibrium trip price condition in the pre-peak period, which is the similar condition for the pre-peak period. (3) traffic volume conservation constraint in the two time periods, which states that the total traffic volume between the origin and the destination in a period must be equal to the sum of the volumes on the toll route and the free route in that period. (4) nonnegativity condition for traffic volumes, that is, the route volumes must also be nonnegative.

The optimized program therefore is to find the optimal traffic volumes on each route (toll route and free route) for each period (peak and pre-peak), which maximize the net benefit subject to the two constraints. The optimal traffic volumes can be solved from the first order conditions for the Lagrangian which constitute a set of necessary conditions. The second-best congestion tolls in the peak and pre-peak periods can then be obtained respectively from the first two constraints by substituting the solution of the optimal traffic volumes in the corresponding constraints. Some initial findings are: (1) the second-best tolls are endogenously determined by the traffic volumes; and (2) the second-best toll on the toll route for each period is equal to the marginal congestion cost plus two adjustment terms. In general it is rather difficult to find the traffic volume analytically, the congestion tolls can only be computed in a numerical way or by simulation. A simulation study of Model SB will be given in the next section.
3.2 Formulation of two alternative congestion pricing problems

In order to evaluate the second-best congestion pricing scheme, it is necessary to study two other regimes for the system: the first-best problem (Model FB) in which congestion tolls can be imposed on both routes in both time periods, and the no-toll problem (Model NT) in which congestion tolls cannot be imposed on either route.

For the first-best problem, there are no pricing constraints (1) and (2). In this case net benefits $B-C$ are maximized, subject to traffic volume conservation constraints (3) and nonnegativity condition (4). The optimality conditions of model FB say that the trip price for each route and for each period is given by its marginal cost. From the optimality conditions, the first-best congestion tolls can be derived analytically in terms of the marginal congestion cost, which represents the cost that the traveler imposes on all other travelers by adding to the level of congestion.

For the no-toll problem, no maximization is involved. Traffic volumes are determined by the equilibrium conditions with traffic volume conservation constraints (3) and nonnegativity condition (4). The equilibrium condition says, in each period, the equilibrium price of a trip is equal to the average cost on both the free route and the toll route.

4 Simulation study of peak and off-peak congestion pricing models

4.1 Simulation procedure for solving the congestion pricing models

This section outlines solution procedures to solve the congestion pricing models presented in the previous section: Model SB, Model FB, and Model NT. In general, it is difficult to solve the three models analytically. Therefore there is a need to conduct a simulation study to solve the models numerically. For each of the three models, the first-order conditions are equivalent to the system of equations for the traffic volume allocation.

A procedure, which uses the Newton’s method to solve the equivalent nonlinear system of equations, is applied to find the solutions for the three models, by specifying a set of reasonable parameter values for the cost and the demand functions and initial values of the traffic volume allocations for each of the three problems. Given input data, the procedure generates one set of solutions for the three models.

The major parameter values for the average cost function, which applies the FHWA function (see Branston, 1976 for a review), include the uncongested travel time (minutes) per mile and the level of capacity (vehicles per hour) for each route. In addition, the value of travel time (cents per minute) (Small, 1982), and the unit operating costs (cents per mile) (Small, 1992) also need to be specified. In order to conduct the simulation study, the uncongested travel time and level of capacity for both the toll route and free route need to be assigned values. The simulation study is designed to solve the models for different cases, which represent different scenarios in the study network (that is, the supply side). A basic criterion is that the toll route has an uncongested flow travel time lower than or equal to the free route. Given this criterion, four cases are considered:

Case 1: The toll route has a lower free flow travel time than the free route; and two routes have identical capacities.
**Case 2:** The toll route has a lower free flow travel time and a larger capacity than the free route. This case represents the scenario in which congestion tolls can be imposed in a major portion of the network.

**Case 3:** The toll route has a lower free flow travel time and a smaller capacity than the free route. This case represents the scenario in which congestion tolls can be imposed only in a small portion of the network.

**Case 4:** This is a special case in which the two routes have identical free flow travel times and capacities.

For Cases 1-4, the two routes are assumed to have the same lengths which are equal to one mile.

The demand functions for the peak and pre-peak periods are assumed to be linear. Some key parameters and the corresponding assumptions include: (1) the potential demand (that is, the demand with zero prices) in the peak period is higher than the pre-peak; (2) negative own-price coefficients and positive cross price coefficients for each period. As implied by the integrability condition, the cross-price effects are equal. It is also assumed that the own-price effects outweigh the cross-price effects. The inverse demand functions are also linear with equal cross effect coefficients. For the base case simulation, the demand parameters are based on Wohl and Hendrickson (1984).

After specifying the cost and demand functions, one can obtain the optimal traffic volume allocations by solving the nonlinear systems of equations for model SB, model FB, and model NT, respectively. The solutions to the three models: second-best (SB), first-best (FB), and no-toll (NT), can be computed numerically by applying the Newton’s method, to each of the four cases considered in the cost functions. For each case, the output results include, (1) optimal traffic volume allocations; (2) congestion tolls, equilibrium average cost, and trip prices; and (3) social welfare, welfare gains, and relative welfare improvement, for models SB, FB and NT, respectively.

**4.2 Analysis results of the simulation study**

The analysis focuses on the efficiency issue of congestion pricing - the effectiveness of the congestion pricing in reallocating traffic volumes to maximize social welfare. Three major issues have been investigated: (1) the impacts of the congestion tolls on the traffic volume allocations; (2) the differences between the second-best and the first-best tolls; and (3) the social welfare properties of the second-best congestion pricing scheme.

The second-best congestion pricing policy has three major impacts on the traffic volume allocations: (1) the diversion of the peak period traffic to the free route, (2) the shift of the peak period traffic to the pre-peak period, and (3) the reduction in total traffic volumes. However, the second-best policy is less effective than the first-best policy in reallocating traffic volumes primarily because the FB scheme allows the tolls on both routes.

The second-best tolls differ sizably from the first-best tolls. In the peak period, the second-best tolls are smaller than the first-best tolls. For example, the peak period toll on the toll route in Case 2 is $1.08 per mile in model FB versus $0.48 per mile in model SB. In the pre-peak period, for most cases, the second-best tolls are smaller than the first-best in some cases. A noticeable result for Case 4 is that the SB toll in the pre-peak period is...
negative, which is equal to -0.8 (cents per mile per vehicle). The negativity implies that under some circumstance (for example, the two routes have equal free flow travel times, etc.), people are encouraged to travel in the pre-peak period by being subsidized instead of being tolled in order to achieve the second-best goal.

The welfare property of the second-best congestion pricing policy is less desirable than the first-best policy. Although both the second-best and the first-best policies have social welfare gains against the no-toll policy, the welfare gains yielded by the second-best are less than half of the welfare gains by the first-best scheme (45.8%, 45.3%, 19%, and 11.4% for Cases 1-4, respectively). Therefore, the failure to impose a congestion toll on a major part of the network may results in a loss of the potential welfare gains.

Other simulations were run (for Case 2) to test the sensitivity of the conclusions to changes in the demand parameters. In two cases, the sensitivities of volumes to prices (the coefficients) were increased (decreased) by one-third. Larger values for the coefficients caused the first-best toll in the peak period to be somewhat lower, but had virtually no effect on other results. The welfare gain in the second-best case is 48.1% of the gain in the first-best case. Similarly, smaller values for the coefficients generated a higher peak period toll in the first-best case and produced a relative welfare gain in the second-best case of only 38.6%. In two other cases the potential demand for trips during the peak period was increased (decreased) by 10%. As one would expect, the increase in demand (decrease in demand) caused the peak period tolls to rise (fall), but the relative welfare gains in the second-best case changed very little.

4.3 Summary of major findings

Overall, the simulations reported may lead to the following conclusions:

1. Within the range of values tried out, overall results for various cost and demand parameters are not sensitive to these changes, but there are some differences worth noting.

2. Cases 2 and 3 show that coverage of system with toll from 0, 1/3, 2/3, 1 generates a nonlinear percentage of potential welfare gains: 0, 19, 45, 100.

3. Cases 1 and 4 show that the system with faster toll route yields higher percentage of welfare gains: 46% vs. 11%.

4. Demand parameters consist of slopes and intercepts. For the assumed slopes, the basic conclusions hold but more empirical studies are needed to obtain estimates of the slopes. For the intercepts it is shown that with the potential peak demand reduced, percentage of possible welfare gains rises; second-best tolls fall in the peak and rise in the off-peak period.

5. There is a possibility of negative second-best toll in the off-peak period with low potential off-peak demand.
5 Extensions of the existing models and further research topics

This study has investigated the second-best congestion pricing scheme for an urban road system in which congestion tolls are not allowed on a major part of the system due to technical and/or political constraints. The current research may be extended by the following topics:

Firstly, the two-period (peak and off-peak) congestion pricing models could be extended to multi-period models, for example, pre-peak, peak, and post-peak periods. In addition, there is no mode choice in the study. Traveler’s mode choice, for example, automobile and transit, could be added to the current model (Huang, 2000).

Secondly, the network used in the study consists of two routes connecting an origin and a destination. Further research could extend the network to a general network with more than two routes. For example, Verhoef (2002) extended the single-period second-best congestion pricing model to a general road network. Liu and Boyce (2002) examined the problem of the multi-period optimal (first-best) congestion pricing for a general transportation network.

Thirdly, the optimization formulations of the welfare-maximizing models (SB and FB) assume symmetry or integrability conditions on the inverse demand functions. The symmetry assumption is restrictive in terms of applications and precludes more realistic modeling of multiple O-D pairs and time periods. As indicated in Nagurney (1999), the line integral problem can be overcome by introducing a more general formulation, the variational inequality (VI) problem, which eliminates the line integral and the associated symmetry assumption. Liu and Boyce (2002) present a variational inequality formulation of the multi-period first-best congestion pricing problem for a general transportation network.

Lastly, the study focuses on the efficiency issue and assumes that travelers are homogeneous, that is, all travelers have the same value of time, and belong to the same income group. Further studies could consider multiple groups of travelers with different values of time (Yang and Zhang, 2002), or with different income levels, for example, rich and poor. With this extension, other issues such as equity could also be examined.

6 References


