EXTERNALITIES, AVERAGE AND MARGINAL COSTS, AND TOLLS ON CONGESTED NETWORKS WITH TIME-VARYING FLOWS

MALACHY CAREY
University of Ulster, N. Ireland, United Kingdom

ASHOK SRINIVASAN
Purdue University, West Lafayette, Indiana

(Received October 1989; revisions received February 1991, July 1992; accepted July 1992)

For congested networks on which flows vary over time, we derive system marginal costs, user perceived costs and user externality costs, for each arc and path. We also obtain a set of optimal congestion tolls and flow controls which may be used to shift the user determined flows toward a socially preferred pattern. An important way in which our results differ from the usual static analysis is that the social cost externality depends not only on the level of congestion, but also on the rate of increase or decrease of congestion. This is intuitively explicable as follows. Consider users delayed on an arc. Their delays will be further compounded or multiplied if congestion has increased during the time they are delayed. On the other hand, their delays will be reduced if congestion has declined during the time they are delayed. This multiplier effect is such that the resultant dynamic externalities can easily be a few times larger, or smaller, than the externalities derived in the usual static analysis. As a result, the congestion tolls or tariffs which are usually proposed or advocated, based on static analysis, may be inappropriate. The results are illustrated with a numerical application to a small network example.

An important part of the total cost of transportation is due to congestion. Congestion imposes travel time costs on all classes of traffic, including urban commuters and industrial and commercial transport. Furthermore, reducing and controlling congestion requires very large annual expenditures on construction and maintenance of roads, etc. Transportation and traffic models are needed to quantify these costs, to understand, manage and control the complex patterns of traffic flows and congestion costs, and to examine the tradeoffs between congestion costs and investment costs.

Existing congested network models are mainly static—that is, they assume that traffic flows are constant over time. Here we adapt and use a model which allows congested traffic flows to vary over time, especially over peak demand periods. We find that this greatly changes the nature and magnitudes of congestion costs, as compared with the results obtained from existing static models.

Each additional user of a congested road or other facility imposes a congestion cost (known as an externality) on other users, by slowing them down, and perhaps by increasing inconvenience, pollution, risk of accidents, etc. However, in deciding whether, when, or by which route to travel each user normally considers only his/her own costs and does not consider the costs which he imposes on others. But by imposing a toll or tariff exactly equal to the externality we can ensure that the users’ optimal private choices will also be optimal social choices. User externalities are important because they can be very large, and can easily be much larger than the cost experienced or perceived by the individual user. Thus, in making a travel choice the user may be taking account of only a small part of the total social cost.

The above ideas are usually discussed in the context of road traffic, hence in the present paper we adopt the same focus. However, the analysis also applies to other types of congested networks for which the individual users of the network, acting separately, seek to minimize their separate individual costs (or maximize their individual benefits) from travel, and in doing this choose their own individual starting times or routes.

Furthermore, the discussion and analysis in this paper is of interest in much more general network contexts because the discussion can be restated in...
terms of average versus marginal costs, in particular, the average and marginal costs of utilizing each facility, department, (arc), etc. Network optimization models assign flows based on marginal costs. However, in practice average cost pricing is prevalent in industry, including transportation, communications, and other network related industries. As a result, average costs, marginal costs and the difference between these is of interest for the purposes of accounting, costing, pricing, marketing, regulation, etc. The difference between average and marginal costs is referred to as the externality in the present paper. We recognize, of course, that this difference is not always an externality in the economic sense (Henderson and Quandt 1980). It is an externality in the economic sense if it is borne by persons other than the decision maker.

The topic of externalities and average versus marginal cost pricing and allocation has not been dealt with in the general literature on dynamic network models. See, for example, surveys by Bookbinder and Sethi (1980) and Aronson (1989). Average versus marginal costs are discussed in the economics literature, the management accounting literature (see, for example, a textbook by Kaplan 1982), and the operations research literature (for example, Miller and Buckman 1987). But this concern has not yet been extended to time-varying network flows.

There is a literature that relates to aspects of the problem for traffic and transportation, but, in each case, some of the important features mentioned above are not included. Thus, in developing and analyzing externalities and optimal congestion tolls, some authors consider a single facility or bottleneck with time-varying flows (e.g., Agnew 1977, Else 1981, Henderson 1974, 1981, Hendrickson and Kocur 1981, and Alfa 1986, whereas others consider a network of facilities, but do not allow time-varying flows (e.g., Walters 1961, Dafermos and Sparrow 1971, Florian 1984, Sheffi 1985). Several authors consider congested time-varying flows, without considering externalities or congestion tolls, but their work may eventually be extended to include externalities and tolls. Thus, single facilities, bottlenecks, or a few competing bottlenecks, are considered by Hurdle (1981), Ben Akiva, Cyna and DePalma (1984), Alfa (1986) and others, and networks of facilities (without route choice) are considered by D'ans and Gazis (1976), and others. Merchant and Nemhauser (1978a, b), Ho (1980), Carey (1987), Carey and Srinivasan (1988), and Friesz et al. (1988) consider time-varying flows on networks, with route choice, but without considering externalities or tolls. It is this work that we use as the starting point in the present paper.

In this paper, we find the behavior of externalities (and, hence, optimal tolls) to be much more complex than the variation of optimal tolls with different levels of congestion. Even if the level of congestion is the same at two times, the externalities (and, hence, the optimal tolls) may be very different at these two times, depending on whether congestion is building-up to a peak or falling-off from a peak. In contrast, static models predict the same externalities at two times of day having the same level of congestion, because static models do not consider the effects of travel choices in one period on congestion in later periods. Static models neglect the intertemporal dimension of the externalities imposed by travelers.

As mentioned, in the case of road traffic it has frequently been proposed that tolls equal to user cost externalities be imposed on users. However, the relevance and importance of the investigation of externalities does not depend on the extent to which (dynamic) congestion tolls are currently implementable in practice, because:

a. The analysis and measurement of externalities is useful in evaluating proposals other than tolls, e.g., investments in capacity expansion, traffic controls, flexible work start time arrangements, etc.

b. Current and foreseeable advances in microprocessors for in-vehicle use can be used to make measuring and collecting dynamic tolls relatively simple and cheap in the foreseeable future.

To provide a benchmark with which to compare our analysis of dynamic (time-varying) flows we first set out, in Section 1, the well known static traffic assignment model, including the corresponding expressions for user externalities and static optimal tolls. Section 2 is devoted to the dynamic model and includes our main results: In subsection 2.1 we set out a model for optimizing congested network flows which vary over time, derive from this the system marginal costs (subsection 2.2), and the costs as perceived by the average or typical user of the network (subsection 2.3). From these marginal and user perceived costs we obtain, in subsection 2.4, the user externalities and optimal dynamic tolls. In Section 3 we show that all the discussion and results from the fixed-demand case (Section 2) carry over to the case of variable price-elastic demands. This is particularly important if we contemplate introducing optimal tolls, because one of the main purposes and effects of such...
tolls is to shift the timing of travel demands, say from peak to off-peak. Section 4 provides further comparisons between the dynamic and static externalities and tolls, and illustrates the relationships with some simple network examples. We conclude (Section 5) with remarks on implementation.

1. STATIC EXTERNALITIES AND TOLLS

In this section, for the sake of later reference and comparison, we briefly set out a standard, well known static traffic assignment model and state the familiar expression for user cost externalities, and hence optimal congestion tolls, for the static case. Let the traffic network be represented by a directed graph \( G = (N, A) \), where \( N = \{1, 2, \ldots, q, \ldots\} \) is the set of nodes and \( A = \{1, 2, \ldots, j, \ldots\} \) is the set of directed arcs joining these nodes. Let \( A(q) = \{j \in A \mid j \text{ points out of node } q\} \) and \( B(q) = \{j \in A \mid j \text{ points into node } q\} \). Let \( d_j \) be the flow rate passing through link \( j \) and let \( f_j(d_j) \) be the time or cost of traversing arc \( j \). Also, \( f_j(d_j) \) is usually assumed to be a positive, convex, non-decreasing function. Let the exogenous travel demands at node \( q \) be \( F_q \).

Then the well known model for optimizing static flows on a congested single destination network is as follows. (We state here a single rather than multiple destination formulation to facilitate comparison with the single destination dynamic model.)

**Problem S**

Minimize \( Z_s = \sum_{j \in A} d_j f_j(d_j) \)

subject to

\[
\sum_{j \in A(q)} d_j = F_q + \sum_{j \in B(q)} d_j \quad \text{for all } q \in N \tag{1}
\]

\( d_j \geq 0 \quad \text{for all } j \in A. \tag{2} \)

Let \( \lambda_q \) be the Lagrangian multipliers associated with (1). The Kuhn-Tucker necessary conditions for optimality include the following. Let arc \( j \) point out of node \( k \) and into node \( q \). Then,

\[
\lambda_k - \lambda_q = f_j(d_j) + d_j f'_j(d_j) \tag{3}
\]

for all utilized arcs (i.e., arcs \( j \in A \) for which \( d_j > 0 \)). Then \( \lambda_k \) is the marginal cost of traveling from the beginning of arc \( j \) to the destination and \( \lambda_q \) is the marginal cost of traveling from the end of arc \( j \) to the destination. Thus, (3) gives us an expression for the social cost of a marginal trip on arc \( j \). It is made up of two components:

1. \( f_j(d_j) \), which is the time or cost experienced/perceived by an individual user traversing arc \( j \); and
2. the static or contemporaneous externality \( e_j = d_j f'_j(d_j) \), which is the cost the marginal user imposes on others traveling on the arc.

Optimal tolls are the additional costs that have to be imposed on each traveler if we wish to ensure that, in deciding when and whether to travel, he takes account of the full social cost of his trip. Thus, the optimal toll for each arc \( j \) is a toll set exactly equal to the user externality \((d_j f'_j(d_j))\).

2. DYNAMIC FLOWS, MARGINAL COSTS, EXTERNALITIES AND TOLLS

Our objective here is to derive an operational measure of user externalities, and thus optimal congestion tolls. Recall that the externality imposed by each user is defined as the amount by which the social marginal cost imposed by the user exceeds (or falls short of) the cost actually experienced or perceived by the user. Thus, to obtain an expression for the externality we first obtain an expression for the social marginal cost (subsection 2.2), and the user perceived cost (subsection 2.3), and then subtract one from the other (subsection 2.4). To obtain the social marginal cost (subsection 2.2) we start (subsection 2.1) from a model which optimizes the system of network flows over time. This is the same model set out in Carey (1987) and Carey and Srinivasan (1988), and is closely related to the work of Merchant and Nemhauser (1978a, b) and Ho (1980). However, none of these authors discusses deriving externalities or optimal tolls from the model, nor do they suggest how this might be done.

To help interpret the dynamic externalities and tolls in subsection 2.4, we introduce some relationships which hold exactly for static models, hold exactly at some point on each arc in the dynamic flow model, and hold approximately in each period in the dynamic model. This allows us to interpret the dynamic externality by decomposing it into the sum of two terms: one term is the well known static externality obtained in static models (Section 1), and the second term is an analogous intertemporal effect which arises only when flows are varying over time.
2.1. Dynamic System Optimizing Model

The traffic network is represented, as before, by a directed graph \( G = (N, A) \). The destination or sink node is denoted by \( n \). Let the overall planning period be divided into \( T \) equal time intervals \( t = 1, \ldots, T \).

Let the exogenous inflow or input at node \( q \) in period \( t \) be \( F_{q,t} \). Let \( x_{j,t} \) be the number of vehicles on arc \( j \) at the beginning of the \( t \)th time period, and let \( d_{j,t} \) be the number of vehicles that are admitted onto arc \( j \) during the \( t \)th time period. Let the travel costs be proportional to time, thus \( kcx_{j,t} \), represents the travel cost incurred by the volume \( x_{j,t} \) on arc \( j \) during period \( t \), where \( c \) represents the length of the time period and \( k \) is the value of time to the average traveler.

Corresponding to each arc in each period there is a capacity function or congestion function \( g_j(x_j) \) that represents the natural (or uncontrolled) rate of outflow from arc \( j \) in period \( t \), and a variable \( v_{j,t} \), that represents the actual (or controlled) rate of outflow from arc \( j \) in period \( t \). Also, we assume that the capacity function \( g_j(x_j) \) is a concave, nonnegative, nondecreasing function, starting from the origin \( g_j(0) = 0 \). (In the usual static assignment model, \( d_j = g_j(x_j) \) and \( \tau_j = f_j(d_j) \), where \( \tau_j \) is the time taken to traverse arc \( j \). Thus one can derive \( g_j(x_j) \) from the usual arc travel time/cost function \( f_j(d_j) \), as follows: (volume on arc) = (flow rate) \cdot (arc traversal time), hence \( x_j = d_j f_j(d_j) = g_j^{-1}(d_j) \). Hence, given that \( f_j(d_j) \) is a positive, convex, nondecreasing function, it can be shown that \( g_j(x_j) \) is a positive, concave, increasing function.)

To ensure that the actual outflow \( v_{j,t} \), from each arc in the network is less than or equal to the capacity outflow \( g_j(x_j) \) from that arc, let

\[
\{a_j | g_j(x_j) \geq v_{j,t} \} \quad \text{for all } j \in A, t = 1, \ldots, T - 1.
\]

(4)

The items in \{\} brackets in (4)–(6) denote Lagrange multipliers, which are used below. Any slack \( s_{j,t} = g_j(x_{j,t}) - v_{j,t} \), in the above constraint will represent the number of travelers who are prevented from leaving arc \( j \) in period \( t \). Even though these users will incur a cost for remaining on the same arc for an additional period, cost savings of a larger magnitude can accrue on succeeding links as congestion may thereby be lower than what it would have been if there were no restriction on the arc outflow. The system costs of such an assignment can therefore be lower than an assignment without such controls. The values of \{\( s_{j,t} \)\} therefore can be interpreted as an optimal pattern of flow controls. Carey (1987) shows that the optimal value of these flow controls will usually be zero (see Appendix A).

The second set of constraints are the flow conservation or stock adjustment equations stating that the volume on an arc in one period equals the volume in the previous period, minus the inflow plus the inflow,

\[
\{\lambda_{j,t} | x_{i,t+1,j} = x_{i,t} - v_{i,t} + d_{i,t} \}
\]

for all \( j \in A, t = 1, \ldots, T - 1 \). (5)

For the artificial arc \( j = d \) which points out of the destination, (5) reduces to \( x_{i,t+1,d} = x_{i,d} + d_{i,d} \) for all \( t = 1, \ldots, T - 1 \), so that \( x_{i,d} \) represents the cumulative volume which has arrived at the destination up to time \( t \).

The third set of constraints states that for each node the inflow to the node equals the outflow from the node,

\[
\{\mu_{q,t} | \sum_{j \in \bar{B}(q)} d_{j,t} = F_{q,t} + \sum_{j \in \bar{B}(q)} v_{j,t} \}
\]

for all \( q \in N, t = 1, \ldots, T - 1 \). (6)

For the destination node \( (q = n) \), (6) reduces to \( d_{d,t} = \sum_{j \in \bar{B}(n)} v_{j,t} \), for \( t = 1, \ldots, T - 1 \).

The fourth set of constraints ensures nonzero initial loadings on the arcs.

\[
x_{i,0} = \bar{x}_{i,0} > 0 \quad \text{for all } j \in A,
\]

(7)

where \( \bar{x}_{i,0} > 0 \) is given. This ensures that the Slater constraint qualification is satisfied, thus ensuring that the Kuhn-Tucker optimality conditions will hold at an optimum (see Carey 1986 for details).

Last, we have nonnegativity constraints on the arc volumes, inflows and outflows,

\[
x_{i,t} \geq 0 \quad \text{for all } j \in A, t = 2, \ldots, T,
\]

(8)

\[
(v_{j,t}, d_{j,t}, F_{q,t}) \geq 0 \quad \text{for all } q \in N, j \in A, t = 1, \ldots, T.
\]

(9)

A model which minimizes system travel costs can now be stated as follows.

Program C

Minimize \( z = \sum_{i=1}^{j} \sum_{t \in A} c^{k}x_{i,t} \),

subject to (4)–(9).

In subsections 2.2–2.4 we derive expressions for dynamic system marginal costs, user perceived costs and hence user externalities. To put these expressions in perspective it is useful to compare them with the corresponding expressions for the static case of Section 1. To facilitate this comparison of the static and dynamic cases we show, in the lemma below, how the arc travel time function \( f(v) \) from the static
model relates to the flow rate function \( g(x) \) of the dynamic model. This lemma will be used in subsections 2.2–2.4.

**Lemma 1.** Let the flow rate \( v_j \), and hence the volume \( x_j \), on arc \( j \) be held constant (static) over time. Let \( f_j(v_j) \) be the time taken to traverse arc \( j \) when the flow rate is \( v_j \) (as in Section 1), and let \( g(x_j) \) be the outflow rate from arc \( j \) when the volume on arc \( j \) is \( x_j \). Then,

\[
x_j/g(x_j) = f_j(v_j),
\]

that is, \( x_j/g(x_j) \) equals the user perceived arc-traversal time.

\[
1/g'_j(x_j) = f_j(v_j) + v_j f'_j(v_j),
\]

that is, \( 1/g'_j(x_j) \) equals \( m_j(v_j) = (f_j(v_j) + v_j f'_j(v_j)) \) the system marginal cost of traversing an arc. Subtracting (10) from (11) gives,

\[
(1/g'_j(x_j) - x_j/g(x_j)) = v_j f'_j(v_j),
\]

which is the well known static externality \( e_j \) (or optimal static congestion toll) measured in travel time units (see Section 1).

**Proof.** Since the flow rate is constant along the arc, \( x_j = v_j f_j(v_j) \) and \( v_j = g(x_j) \), and substituting the latter into the former yields (10). Also, differentiating \( x_j = v_j f_j(v_j) \) and \( v_j = g(x_j) \), respectively, yields \( dx_j/dv_j = f_j(v_j) + v_j f'_j(v_j) \) and \( dv_j/dx_j = g'_j(x_j) \). But \( dx_j/dv_j = 1/(dv_j/dx_j) \), hence (11) follows. Subtracting (10) from (11) gives (12), which completes the proof.

### 2.2. System Marginal Costs

The cost imposed on the system by a marginal user entering an arc can be obtained from the Kuhn-Tucker conditions, which necessarily hold at an optimum of program \( C \), and which can be set out as follows. Let \( \{a_i\}, \{\lambda_j\} \) and \( \{\mu_{ij}\} \) be the Lagrangian multipliers associated with (4), (5), and (6), respectively. Let \( g'_j(x_j) \) denote \( dg(x_j)/dx_j \). The Kuhn-Tucker conditions (Bazaraa and Shetty 1979) consist of (4)-(9), together with,

\[
\begin{align*}
\lambda_{i-1,j} - \mu_{ij} &\geq 0, \quad (d_{ij} \geq 0), \\
k - \lambda_{i-1,j} - a_{ij} g'_j(x_j) + \lambda_j &\geq 0, \quad (x_j \geq 0),
\end{align*}
\]

for all \( j \in A \) \quad (13a)

\[
k - \lambda_{i-1,j} \geq 0, \quad (x_Tj \geq 0),
\]

for all \( j \in A \) \quad (13b)

\[
a_{ij} - \lambda_j + \mu_{ij} \geq 0, \quad (v_{ij} \geq 0),
\]

for all \( j \in B(q), \ q \in N \) \quad (14)

\[
\lambda_{ij} - \mu_{iq} \geq 0, \quad (d_{ij} \geq 0),
\]

for all \( j \in A(q) \) \quad (15)

\[
a_{ij} \geq 0, \quad (g(x_j) \geq v_{ij}),
\]

for all \( j \in A \) \quad (16)

for all periods \( t = 1, \ldots, T - 1 \), and complementary slackness conditions between the pairs of inequalities in (13a)-(16).

The Lagrange multipliers \( \mu_{ij} \) and \( \lambda_j \) can be interpreted as follows. Here \( \mu_{ij} \) is the Lagrange multiplier associated with (6) and hence can be interpreted as the marginal system cost of increasing the exogenous inflow \( F_{iq} \) at node \( q \). In other words, \( \mu_{iq} \) is the marginal cost of an additional user entering at node \( q \) and traveling from there to the destination. Also, \( \lambda_j \) can be interpreted in various ways, however, the interpretation which is most useful here is derived as follows. We assume that the inflow \( d_{ij} \) to arc \( j \) is nonzero (\( d_{ij} > 0 \)). Then from the above Kuhn-Tucker conditions \( \lambda_{ij} = \mu_{iq}, \ j \in A(q) \), hence \( \lambda_{ij} \) has the same interpretation as \( \mu_{iq}, \ j \in A(q) \). That is, \( \lambda_{ij} \) is the marginal cost of an additional user entering arc \( j \) (at node \( q \)) and traveling from there to the destination. Note that an equivalent interpretation of \( \lambda_{ij} \) above is the marginal savings due to a user exiting at node \( q \). Hence \( (\lambda_{ij} - \mu_{iq}), j \in B(q) \), can be interpreted as the marginal cost of using arc \( j \) in period \( t \): \( \lambda_{ij} \) is the marginal cost of an additional user entering the arc, and \( \mu_{iq}, j \in B(q) \) is the marginal savings from a user exiting the system at the end of the arc. The following relationships between the Lagrange multipliers have useful interpretations and analogies with the static case.

**Proposition 1.** When the outflow \( v_{ij} \) from arc \( j \) is positive \((v_{ij} > 0)\) and the optimal flow control for arc \( j \) is zero (i.e., \( g(x_j) = v_{ij} \); see Appendix A), then the marginal cost of traversing from the beginning of arc \( j \), at time \( t - 1 \), to the destination via any utilized path is,

\[
\lambda_{i-1,j} = kc + \lambda_j (1 - g'_j(x_j)) + \mu_{iq} g'_j(x_j)
\]

and the marginal cost of using arc \( j \) in period \( t \) is,

\[
(\lambda_{ij} - \mu_{iq}) = kc/g'_j(x_j) + (\lambda_{ij} - \lambda_{i-1,j})/g'_j(x_j).
\]

**Proof.** To obtain (17) proceed as follows. If \( v_{ij} > 0 \), then by complementarity, the first inequality in (14) is a strict equality. Also, from (4), \( v_{ij} > 0 \) implies that \( g'_j(x_j) > 0 \), hence (by the assumed properties of \( g(\cdot) \)), \( x_j > 0 \), and by complementarity, the first inequality in (13) is a strict equality. Now using the equality
form of (14) to substitute for $a_{ij}$ in the equality form of (13) yields (17). To obtain (18), multiply through (17) by $1/g_j'(x_{ij})$ and rearrange.

Using Lemma 1 we can obtain the following interesting approximation and interpretation for (18). From (11) in Lemma 1, $1/g_j'(x_{ij})$ approximates $m_j(v_{ij})$ the marginal cost of traversing arc $j$. Multiplying by $k$ and $c$, brings this to money units per period. Then substituting for $1/g_j'(x_{ij})$ in (18) gives,

$$\lambda_{ij} - \mu_{ij} = kc m_j(x_{ij}) + \Delta \lambda_{ij} m_j(x_{ij})$$

$$= (kc + \Delta \lambda_{ij}) m_j(x_{ij}).$$  \hspace{1cm} (19)

The first term on the right-hand side of (19) is identical to the usual static marginal cost of traversing an arc. This term is always positive because $k$, $c$, and $m_j(*)$ are positive. The second term on the right-hand side of (19) is an intertemporal cost effect, and its sign depends on whether $\lambda_{ij}$ is increasing ($\Delta \lambda_{ij} > 0$) or decreasing ($\Delta \lambda_{ij} < 0$); $\lambda_{ij}$ tends to increase in prepeak periods and decrease in postpeak periods.

2.3. Costs as Perceived by Individual Users

If all users of a traffic network are assumed to be treated the same, they will all experience the same actual or expected travel times. Thus each user experiences the average travel time, that is, the sum of all users' travel times divided by the number of users. This average travel time experienced or perceived by individual travelers can, of course, differ substantially from the marginal travel cost or time with which a system optimizer is concerned. Let,

- $\lambda_{ij} = \text{the user perceived travel time or cost from the beginning of arc } j \text{ to the destination;}$
- $\mu_{ij} = \text{the user perceived travel time or cost from the end of arc } j \text{ (node } q) \text{ to the destination.}$

From these two definitions it follows that $(\lambda_{ij}'' - \mu_{ij}''') = \text{the user perceived cost of traversing arc } j \text{ at time } t.$

Proposition 2. The user perceived (average) travel cost, from the beginning of arc $j$ at time $t = 1$ to the destination, can be approximated by,

$$\lambda_{ij}'' = kc + \lambda_{ij}''(1 - \theta_{ij}) + \mu_{ij}'' \theta_{ij},$$  \hspace{1cm} (20)

where $\theta_{ij} = g_j(x_{ij})/x_{ij}$. Dividing through by $\theta_{ij}$ and rearranging gives,

$$\lambda_{ij}'' = kc/\theta_{ij} + (\lambda_{ij}'' - \lambda_{ij}'' '- \theta_{ij})$$

as the user perceived (average) cost of traversing arc $j$ at time $t$.

Remarks on (20)

a. Note the similarity of (20) to (17). The only formal difference is that the marginal outflow rates $g_j'(x_{ij})$ are replaced by the average outflow rates $\theta_{ij} = g_j(x_{ij})/x_{ij}$. 

b. There are a number of ways to derive the approximation equation (20). One is set out in the proof/derivation below. Equation (20) is an approximation, due to dealing in discrete time periods rather than continuous flows.

c. An intuitive way to remember (20) is to note that of the volume $x_{ij}$ on arc $j$ in period $t$ the fraction which exists in period $t$ is $\theta_{ij} = g_j(x_{ij})/x_{ij}$, hence the fraction which remains is $(1 - \theta_{ij})$. The corresponding perceived/average travel time for these two groups is $\lambda_{ij}'(1 - \theta_{ij})$ and $\mu_{ij}'$, hence the average over these two groups is $\lambda_{ij}''(1 - \theta_{ij}) + \mu_{ij}'' \theta_{ij}$. Also, traffic which enters arc $j$ in period $t = 1$ incurs a time/cost $kc$ on the arc until period $t$. The sum of these costs yields (20).

Proof/Derivation. Let $r_{ij}$ be the perceived or average time taken to transverse arc $j$ if entering arc $j$ at time $t$. Now consider traffic entering arc $j$ at time $(t - r_{ij})$. The perceived/average cost of getting from the beginning of link $j$ at time $(t - r_{ij})$ to the final destination is $\lambda_{ij}'(t)$. This can be broken down into two components:

a. The cost of traversing arc $j$. This takes $r_{ij}$, periods at a cost of $(kc)$ per period, hence this costs $(kc)r_{ij}.$

b. The cost of getting from the exit of arc $j$ to the destination. By definition this costs $\mu_{ij}'$.

Thus we have,

$$\lambda_{ij}'(t) = (kc)r_{ij} + \mu_{ij}'$$  \hspace{1cm} (22)

Now consider $\lambda_{ij}', t = 1, \ldots, T$ as discrete points on a continuous function $\lambda_j'(t)$ varying over time. A local linear approximation to the time path of $\lambda_j'(t)$ at time $t$ is given by

$$\lambda_j'(t) - r(t)(d \lambda_j'(t)/dt) = \lambda_j'(t - r(t))$$

or in discrete terms,

$$\lambda_{ij}' - r_{ij}(\Delta \lambda_{ij}/\Delta t) = \lambda_{ij}' - r_{ij}.$$  \hspace{1cm} (23)

Let $\Delta t = 1$, so that $\Delta \lambda_{ij}' = (\lambda_{ij}' - \lambda_{ij}' - r_{ij}).$ This reduces (23) to

$$\lambda_{ij}' - r_{ij}(\lambda_{ij}' - \lambda_{ij}' - r_{ij}) = \lambda_{ij}' - r_{ij}$$

and using this to substitute for $\lambda_{ij}' - r_{ij}$ in (22) gives,

$$\lambda_{ij}' - r_{ij}(\lambda_{ij}' - \lambda_{ij}' - r_{ij}) = (kc)r_{ij} + \mu_{ij}'$$
and rearranging gives,
\[ \lambda_{i,j}^{n_i} = (kc) + \lambda_{i,j}^{u_j}(1 - (1/r_{ij})) + \mu_{u_j}(1/r_{ij}). \] (24)

A continuous approximation to the arc trip time \( r_{ij} \) at time \( t \) is given by \( r_{ij} \approx (\text{volume on arc } j \text{ at time } t)/(\text{outflow rate from } j \text{ at time } t) = x_{ij}/g_{ij}(x_{ij}). \) This local approximation is better the shorter the arc and the time period. Substituting \( r_{ij} = x_{ij}/g_{ij}(x_{ij}) \) in (24) gives (20).

Using Lemma 1 we can obtain the following instructive approximation and interpretation for (21). Introducing a time subscript \( t \) on \( x_{ij} \) and \( x_i \) in Lemma 1 gives \( x_{ij}/g_{ij}(x_{ij}) = f_s(v_{ij}) \), for the static case. But \( x_{ij}/g_{ij}(x_{ij}) = 1/\theta_{ij} \), hence \( f_s(v_{ij}) = 1/\theta_{ij} \), and substituting this for \( 1/\theta_{ij} \) in (21) gives,
\[ (\lambda_{ij}^{u_j} - \mu_{u_j}) \approx (kc) f_s(v_{ij}) + \Delta \lambda_{ij}^{u_j} f_s(v_{ij}). \] (25)

Note that \( f_s(v_{ij}) \) is measured in the number of time periods and \( c \) is the length (in say minutes) of each period, hence \( f_s(v_{ij}) \) is in the same time units as \( c \). The first term on the right-hand side of (25) is identical to the usual user perceived cost of traversing an arc. The second term on the right-hand side of (25) is not found in static user equilibrium models. It is an intertemporal effect, and its sign is of \( \Delta \lambda_{ij}^{u_j} \), which is determined by whether \( \lambda_{ij}^{u_j} \) is increasing or decreasing over time; \( \lambda_{ij}^{u_j} \) will tend to increase during prepeak periods and decrease over postpeak periods.

(The second term (\( \Delta \lambda_{ij}^{u_j} f_s(v_{ij}) \)) on the right-hand side of (25) can be interpreted as a local linear (first-order) approximation as follows. From the definitions of \( \lambda_{ij}^{u_j} \), and \( \mu_{u_j} \), we see that the user perceived cost of traversing arc \( j \) may be written as \( (\lambda_{ij}^{u_j} - \mu_{u_j}) = kc f_s(v_{ij}), \) where \( \tau \) is the time at which users must set out on arc \( j \) to traverse it by time \( t \). Thus \( (t - \tau) = f_s(v_{ij}) \). Also,
\[ (\lambda_{ij}^{u_j} - \mu_{u_j}) \approx ((\lambda_{ij}^{u_j} - \lambda_{ij}^{u_j-1})(t - \tau)) \] (25)
\[ = \Delta \lambda_{ij}^{u_j}(t - \tau) \approx \Delta \lambda_{ij}^{u_j} f_s(v_{ij}). \]

Thus we have \( \lambda_{ij}^{u_j} \approx \lambda_{ij}^{u_j} + \Delta \lambda_{ij}^{u_j} f_s(v_{ij}). \) Substituting this for \( \lambda_{ij}^{u_j} \) in (25) gives (20). It is instructive to compare the user perceived costs in (20) and (21) with the corresponding system marginal costs, (17) and (18), respectively. In (17) and (18), \( g'_s(x_{ij}) \) can be interpreted as the fraction of marginal users leaving arc \( j \) in period \( t \), and \( 1 - g'_s(x_{ij}) \) can therefor be interpreted as the fraction of marginal users remaining. Thus (17) and (18) can be interpreted analogously to (20) and (21), but with marginal times/costs substituted for average times/costs.

Finally, note that we can construct alternative derivations and interpretations of the above user perceived costs, and still obtain the same expressions (20) and (21).

2.4. Dynamic Externalities

Subtracting the user perceived arc-traversal cost (21) from the system marginal costs (18) for each arc immediately yields the user’s congestion externality, and hence the optimal congestion toll, for each user of the arc. Thus:

Proposition 3. If the optimal flow controls are zero (see Appendix A), then the user’s congestion externality (and hence the optimal congestion toll) for arc \( j \) is,
\[ E_{ij} = kcE'_{ij} + E'_{ij}, \] (26)
where
\[ E'_{ij} = (1/g'_s(x_{ij}) - x_{ij}/g_s(x_{ij})) \] (27)
and
\[ E''_{ij} = (\lambda_{ij} - \lambda_{ij-1})/g'_s(x_{ij}) - (\lambda''_{ij} - \lambda''_{ij-1})/\theta_{ij} \] (28)
and \( \theta_{ij} = g(x_{ij})/x_{ij} \).

Remark. \( E_{ij} \) is a static or contemporaneous component of \( E_{ij} \), and \( E'_{ij} \) is an intertemporal component. The coefficient \( kc \) merely changes the units of measurement, to account for the time period length \( c \) and money units instead of time units \( k \).

Note that the above expression for \( E'_{ij} \) can be rewritten as \( E'_{ij} = \Delta \lambda_{ij}/g'_s(x_{ij}) - \Delta \lambda''_{ij}/\theta_{ij} \) or as \( E''_{ij} = \Delta \lambda_{ij} E'_{ij} + (\Delta \lambda_{ij} - \Delta \lambda''_{ij})/\theta_{ij} \).

Proof. This follows immediately on subtracting (21), the cost of traversing arc \( j \) as perceived by a user, from (18), the cost imposed on the system by a marginal user.

We can provide a useful and instructive approximation and interpretation for the dynamic externality above by using the relationships from Lemma 1. These relationships: a) hold exactly for each arc when flow rates are constant, b) hold exactly at some point on each arc when flow rates are varying, and c) hold approximately in the dynamic (time-varying flows) case. We use Lemma 1 to approximate and interpret the dynamic externality \( E_{ij} \) as follows.

Proposition 4. The dynamic externality \( E_{ij} \) in (26) can be approximated by \( E_{ij} \approx kcE'_{ij} + E''_{ij} \), where \( E''_{ij} = v_i f'_s(v_{ij}) \) is a static or contemporaneous externality.
identical to the well known static externality in Section 1 and Lemma 1.

\[ e_i^t = (\Delta \lambda_{ij} e_i^j + (\Delta \lambda_{ij} - \Delta \lambda_{ij}^0) f_j(v_{ij})) \]

is an intertemporal effect which is not found in static models.

**Remark.** The static component \( e_i^t \) is always positive, but the intertemporal component \( e_i^t \) can be positive, zero or negative (see Propositions 7 and 8 in Appendix B).

**Proof.** We introduce a time subscript \( t \) for all variables in Lemma 1, so that the relationships (10)–(12) hold for all periods \( t \). These relationships are exact if \( v_{ij} \) and \( x_i \) are constant over time, but even if \( v_{ij} \) and \( x_i \) vary over time the relationships can be treated as approximations, especially if \( v_{ij} \) and \( v_i \) and \( x_i \) are treated as corresponding to the midpoint, or other appropriate interpolation point, in each period. The proposition follows immediately on substituting (10) and (12), with added \( t \) subscripts, into (26).

3. INTRODUCING PRICE ELASTIC TRAVEL DEMANDS

Throughout Section 2 the travel demands \( F_q \) were treated as fixed. Here we allow travel demands to be variables, and show that the discussion and results in Section 2 continue to hold.

We can introduce variable price elastic travel demands in a manner similar to the way they are introduced in static traffic assignment models, and in Carey (1987). If we simply let the \( F_q \)'s be variable in program C, then the \( F_q \)'s will all go to zero because that minimizes the objective function (cost) of C. We therefore let \( F_q \) be variable and introduce into the objective function of program C an expression representing the benefits associated with \( F_q \). The benefit from \( F_q \) may be measured by the integral of the inverse demand function for \( F_q \). This travel demand function may be written as \( F_q = F_q(p_q) \), that is, if the price or cost incurred/experienced by a unit of flow getting from node \( q \) to the destination is \( p_q \), then the travel demand from \( q \) to the destination is \( F_q \). The inverse of this demand function is \( p_q = p_q(F_q) \), and a generalized measure of benefit from \( F = [F_q] \) is the integral,

\[ \int_0^T \sum_{q \in N} \int_0^T p_q(F_q) \ dF_q. \]

More generally, allowing for cross-price effects, the inverse demand system can be rewritten as \( p_q = p_q(F) \).

where \( F = \{F_{i1}, \ldots, F_{iq}, \ldots\} \), in which case the line integral benefit function is

\[ \int_0^T \sum_{i=1}^T \sum_{q \in N} p_q(F_q) \ dF_q. \]

Alternatively, we can represent travel demands and the benefits from travel by introducing a constraint \( q = \sum_{i=1}^T F_{iq} \), where \( F_q \) is the aggregate demand at node \( q \) over the time span \( t = 1 \) to \( t = T \). Let the inverse demand function associated with \( F_q \) be \( p_q = p_q(F_q) \), hence the line integral benefit function is

\[ \sum_{q \in N} \int_0^T p_q(F_q) \ dF_q. \]

**Proposition 5.** All the propositions and discussion for the fixed demand case continue to hold for the variable demand case when the variable demands are introduced in any of the ways outlined above.

**Proof.** The discussion and results in subsections 2.1–2.4 are based on the form and interpretation of the Kuhn-Tucker conditions (4)–(9) and (13)–(16) from the fixed demands program C. None of these K-T conditions, nor their interpretations, are altered by letting the demands \( F_q \) be variable and introducing any benefit function, as above. In particular, in the program with variable \( F_q \)'s all constraints are still linear, hence the K-T conditions necessarily hold at any solution of the program. Also, for each variable demand program in the proposition the K-T conditions consist of:

a. the K-T conditions (4)–(9) and (13)–(16) from the fixed demand program C, and
b. \( \mu_{iq} \) = the inverse demand function for \( t_q \), i.e., \( p_q(F_q) \), or \( p_q(F) \), etc.

The K-T conditions, b, are consistent with (and strengthen) the interpretation of \( \mu_{iq} \), as given in the remark in Proposition 1. The K-T conditions, a, ensure that all of the K-T conditions of Section 2 continue to hold.

4. COMPARING THE INTERTEMPORAL AND STATIC EXTERNALITIES

As shown in Propositions 3 and 4, the marginal traveler entering arc \( j \) at time \( t \) imposes a dynamic externality \( E_{ij} = E_{ij}^t + E_{ij}^t \), where \( E_{ij}^t \) can be interpreted as a static or contemporaneous externality and \( E_{ij}^t \) can be interpreted as an intertemporal externality. The intertemporal externality \( E_{ij}^t \) can be negative or posi-
tive, hence the dynamic externality $E_{ij}$ can be larger or smaller than the static externality $E_{ij}'$.

**Proposition 6.** If arc $j$ at time $t$ is not oversaturated (i.e., if $g_j'(x_{ij}) > 0$, or $f_j'(d_{ij}) > 0$), then the static externality $E_{ij}$ is strictly positive.

**Remark.** The intertemporal externality $E_{ij}'$ can be written as the sum of two components, as in the remark in Proposition 3. The first component is $\Delta \lambda_{ij} E_{ij}$. Since $k > 0$ and $c_i > 0$, the proposition $(E_{ij} > 0)$ ensures that the sign of this component is the sign of $\Delta \lambda_{ij}$. The second component of $E_{ij}'$ is $(\Delta \lambda_{ij} - \Delta \lambda_{ij}'')/\theta_{ij}$, which is positive, negative or zero as $\Delta \lambda_{ij}$ is $>$, $<$ or $= \Delta \lambda_{ij}''$ (see Proposition 8 in Appendix B).

**Proof.** By definition, $E_{ij}' = (1/g_j'(x_{ij}) - x_{ij}/g_j(x_{ij}))$. It is easy to check (e.g., draw a simple diagram) that if a concave function $g(x) > 0$ passes through the origin and is nondecreasing (i.e., $g'(\cdot) > 0$), then the slope $g'(x)$ at any point $x$ is less than the slope $(g(x)/x)$ of the chord to that point, i.e., $g'(x) < g(x)/x$. Inverting gives $1/g(x) > x/g'(x)$, hence $E_{ij}' = (1/g_j'(x_{ij}) - x_{ij}/g_j(x_{ij})) > 0$.

In the usual static analysis (Section 1) the externality depends only on the flow or volume on the arc, but here the externality (26) clearly depends not only on the arc volume $x$ but also on the rates $\Delta \lambda$ and $\Delta \lambda''$ at which $\lambda$ and $\lambda''$ are increasing or decreasing. This is illustrated in Figure 1. The upper part of the figure shows how $\lambda_i$ varies over a hypothetical rush hour or peak period, and the lower part shows how the static externality $E_j$ and the dynamic externality $E_j$ vary over this same peak period.

**Case 1.** $(E_{ij}' > 0 \Rightarrow E_{ij} > E_{ij}')$ From time periods $t_1$ to $t_2$ the dynamic externality is greater than the static externality.

**Case 2.** $(E_{ij}' = 0 \Rightarrow E_{ij} = E_{ij}')$ When congestion levels off at a peak ($\Delta \lambda_{ij} = 0 = \Delta \lambda_{ij}'$), the dynamic externality is equal to the static externality (time $t_2$).

**Case 3.** $(E_{ij}' < 0 \Rightarrow E_{ij} < E_{ij}')$ From time periods $t_2$ to $t_3$, congestion is falling-off and the dynamic externality is less than the static externality.

### 4.1. A Numerical Example

To illustrate the discussion above, we consider the network shown in Figure 2, where nodes 1–6 represent origins and node 7 is the destination. We consider a time span of 30 time periods and assume that for each of the nodes 1–6 the time pattern of origin-destination demands $F_{ij}$ is,

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{ij}$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21–30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>21–30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We assume that all six arcs have the same arc capacity function $v_j = g(x_j)$, which we construct as follows. Let the arc-traversal time be given by a function of the form,

$$ l_{ij} = f(v_{ij}) = t_0(1 + 0.15(v_{ij}/c_i)^4) $$

which is widely used in traffic flow studies, where $l_{ij}$ is the time taken to traverse arc $j$ for travelers entering in time period $t$, $t_0$ is the free flow travel time, $c_i$ is the capacity of the arc, and $v_{ij}$ is the flow rate from the arc in period $t$. The capacity function $g(x_j)$ is derived from the travel time function (29) as follows. As an approximation,

$$ x_{ij} = v_{ij}f(v_{ij}) = t_0(v_{ij} + 0.15 (v_{ij}/c_i)^4)). $$

The inverse of (30) defines the capacity function $v_{ij} = g(x_{ij})$, but (30) is a fifth order polynomial in $v_{ij}$ and does not have an explicit analytic inverse. However, by solving (30) for various values of $x_{ij}$, we can generate points on the inverse, which is all that we need to solve program C. This is very convenient here because it allows us to use (30) to construct a piecewise linear approximation to $v = g(x_j)$, the inverse of (30). We can then use a linear programming package to solve program C. In the example below we used six pieces, each of width 10.0 units, to construct a piecewise approximation to (30). Thus, the capacity function $g(x_j)$ for each arc $j = 1, \ldots, 6$, is defined by passing through the origin ($g_j(0) = 0$) and having the following piecewise linear slopes:

<table>
<thead>
<tr>
<th>$x_j$</th>
<th>0.0–10.0</th>
<th>10.0–20.0</th>
<th>20.0–30.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_j'(x_j)$</td>
<td>0.9740</td>
<td>0.6701</td>
<td>0.3711</td>
</tr>
<tr>
<td></td>
<td>0.2434</td>
<td>0.1810</td>
<td>0.1448</td>
</tr>
</tbody>
</table>

When $g_j(x_j)$ is given by this piecewise linear function, program C reduces to a linear programming problem. Then using the above data we solved program C using a linear programming package: We used the linear programming option in the MINOS (Murtagh and Saunders 1977, 1982). The solution of model C is illustrated in Figures 3 and 4. Figure 3 depicts the volumes on arc 6 of the network, along
with the inflows onto the arc, and the marginal cost of traversing the arc in each period for the dynamic model. We can see from Figure 3 that the marginal costs of traversing the path start falling a few periods before the inflows onto the arc start declining. The inflows onto the arc start to fall a few periods before the downturn in the volume on the arc. The period-by-period dynamic externality and contemporaneous externality on arc 6 are shown in Figure 4. As expected, the intertemporal externality continues to be
greater than the static externality from periods 3–11, corresponding to increasing marginal cost (rising congestion). The marginal cost rises to a peak in period 11 and then starts declining, with the result that the static externality is greater than the intertemporal externality after period 12 (Figure 3).

4.2. Approximating Dynamic Externalities from Static Models

In the previous section, we obtained dynamic externalities using a dynamic assignment model. In (26) we see that the dynamic externality is computed using \( x_{ij}, g(x_{ij}), g'(x_{ij}), \lambda_{ij} \), and \( \lambda_{i-1,j} \). But even if we do not run the dynamic assignment model to obtain the values of these variables, we may still be able to approximate the dynamic externality by obtaining estimates of these variables from some other source and then plugging them into (26). In this way, dynamic externalities may be estimated even from static models.

For example, instead of a dynamic model with a time-horizon of \( T \) periods, one could run \( T \)-static models (as in Section 1), one for each period and estimate the dynamic externalities \( E_{ij} \), as follows. Obtain the optimal values of \( u_{ij} \) and \( \lambda_{ij} \) from each of the \( T \)-static models. Then, following Lemma 1, substitute these \( u_{ij} \)'s in (10)-(12) to obtain estimates of \( x_{ij}/g(x_{ij}) \) and \( 1/g'(x_{ij}) \). Substitute these estimates (and the optimal \( \lambda_{ij} \)'s from the \( T \)-static models) into (27) and (28) to obtain estimates of \( E_{ij} \) and \( E_{ij}' \). Now compute \( E_{ij} \) from (26). The dynamic externality obtained in this way differs from the externalities obtained directly from the dynamic model. This procedure has the advantage of using a well known and widely used static model, but, of course, suffers from the disadvantage that the sequence of static models does not take account of intertemporal costs and choices.

To illustrate this static approximation approach, we implemented it for the same network and used the same data as in the previous section. Note that in the static flow model congestion is represented by travel time functions \( f(u) \), rather than the capacity functions \( g(x) \) used in the dynamic flow model. However, the \( f(u) \) functions that correspond to the \( g(x) \) functions are easy to obtain, as discussed at the beginning of the

![Figure 3. Volume on, inflow onto, and marginal cost of traversing for arc 6.](image-url)
previous section. Static externalities were obtained by running 20 static models, one for each period in which there was a nonzero demand, and dynamic externalities were obtained from the static externalities for each period by using (26). The static externalities and the implied dynamic externalities obtained are shown in Figure 5. The dynamic externalities obtained in this fashion depend on whether congestion is building-up or falling-off.

5. REMARKS ON IMPLEMENTING DYNAMIC TOLLS

In this paper, we have derived optimal tolls that take account of travel demands which vary over time. A distinguishing feature of optimal dynamic tolls, as compared with optimal static tolls, is that the former depend not just on the level of congestion but also on whether congestion is building-up to, or falling-off from a peak.

We conclude the paper with a brief discussion on the implementation aspects of dynamic tolls. Implementing time-varying tolls might prove to be administratively difficult in the short term, but with the advances in microprocessors, in-vehicle microcomputers, etc., it is likely to be feasible in the near future. The paper shifts the focus from administering tolls based on the level of congestion, i.e., the traditional peak and off-peak tolls, to tolls that are also based on whether congestion is rising, remaining the same, or falling-off. A simple second-best scheme to administer such tolls is to charge tolls on a multipart pricing schedule: With a three-part pricing schedule we would have one toll τ₁ for the period when congestion (and, hence, the optimal dynamic toll) is rising, a lower toll τ₂ when congestion is declining, and a third toll τ₃ for the flat peak period.

The main purpose of congestion tolls is to ensure that in deciding when or whether to travel, or by which route, travelers take account of the congestion costs that their travel choice will impose on others. If tolls are to fulfill this price signaling function, then it is essential that the tolls be known accurately to each traveler before he/she makes the corresponding travel choice. For this reason, and for administrative cost reasons, it may be desirable to maintain the daily and weekly pattern of tolls unchanged for months or years at a time. This would still allow different tolls for weekdays, Saturdays and Sundays.

An alternative use of the model would be to compute new tolls every day: In this case, to ensure that these tolls continue to fulfill their price signaling function it would be necessary to display the varying current tolls at key points of the network, or to transmit them to electronic displays in each vehicle. In this case, the model can be rerun, say every thirty minutes,
APPENDIX A

Conditions for Optimal Flow Controls to be Zero

Flow controls for model C are defined just after (4). The results in later sections depend on the optimal value of these flow controls being zero (see Propositions 1, 3, and 5). Sufficient conditions for the optimal values of the flow controls to all be zero are derived and set out in Carey (1987).

These conditions are quite intuitive, and very likely to be satisfied in practice (Carey 1987). The first condition states that the congestion function \( g_i(\cdot) \) should not slope downwards. A downward sloping \( g_i(\cdot) \) would indicate that the arc was “oversaturated” (Carey 1987, p. 64). This is analogous to the well known nonsaturation condition in the case of static network flows—i.e., the condition that the arc capacity function or trip time function \( f_j(\cdot) \) should not bend backwards. The second condition relates to costs, and is satisfied if, for example, the travel cost on each arc is proportional to the time spent on the arc. This condition is slightly stronger for the final period \( T \). However, this too will normally be satisfied in practice because the unit costs on each arc in period \( T \) include the costs of getting from that arc to the destination (Carey 1987, pp. 63–64).

APPENDIX B

The following two propositions are quoted in a remark in subsection 2.4.

Proposition 7. \( \lambda_{ij} > \lambda_{ij}^u > 0 \) for all arcs \( j \in A, t = 1, \ldots, T \) on utilized paths.

Proof. We show this by induction as follows. Consider any period \( t \) and any utilized arc \( j \) pointing out of node \( k \) and into node \( q \). We will show that if

\[ \lambda_{ij} \geq \lambda_{ij}^u \text{ and } \mu_{ij} \equiv \mu_{ij}^u, \]

then

\[ \lambda_{i-1,j} > \lambda_{i-1,j}^u \text{ and } \mu_{ik} \equiv \mu_{ik}^u. \]

Also, on any utilized path \( d_{ij} > 0 \), hence (15) becomes \( \lambda_{ij}' = \mu_{ik} \), where arc \( j' \) is an arc pointing into node \( k \). Also, by definition \( \lambda_{ij}' = \mu_{ik}^u \). Substituting these last two equations into the right-hand equality in (B.2)
\[ \lambda_{i,j} \geq \lambda_{u}^{0}. \]  
\[ \text{(B.3)} \]

Applying (B.1) \[ \Rightarrow \] (B.2) recursively back along any utilized path from the destination yields the first part of the proposition.

We still have to deal with time-space paths which terminate in period \( T \). For these paths we need to show that for the terminal period \( T \),

\[ \lambda_{T-i,j} > \lambda_{T-i-1,j} \quad \text{for all} \quad j \]
\[ \text{(B.4)} \]

and for all other periods \( t < T \),

\[ \mu_{id} \geq \mu_{id}^{u} \quad \text{for all} \quad t \]
\[ \text{(B.5)} \]

for the destination node \( d \). This will complete the proof by induction.

We now show that (B.1) \[ \Rightarrow \] (B.2), as required above.

Subtracting (20) from (17),

\[ \lambda_{i-1,j} - \lambda_{i,j}^{u} = [\lambda_{i,j}(1 - g'_{i,j}) + \mu_{i,j} g'_{i,j}] - [\lambda_{i,j}(1 - \theta_{i,j}) + \mu_{i,j}^{u} \theta_{i,j}]. \]
\[ \text{(B.6)} \]

But,

a. \( 0 \geq (\mu_{i,j} - \lambda_{i,j}) \) for utilized arc \( j \) (from complementarity (14) is an equality, and \( \alpha_{i} \geq 0 \), hence (14) reduces to \((-\lambda_{i,j} + \mu_{i,j}) \leq 0\)).

b. \( g'_{i,j} < \theta_{i,j} \) (see the proof of Proposition 6) hence

\[ (1 - g'_{i,j}) > (1 - \theta_{i,j}). \]

Multiplying through b by \( (\mu_{i,j} - \lambda_{i,j}) \leq 0 \) from a and adding \( \lambda_{i,j} \) to each side gives

\[ (\lambda_{i,j}(1 - g'_{i,j}) + \mu_{i,j} g'_{i,j}) > (\lambda_{i,j}(1 - \theta_{i,j}) + \mu_{i,j} \theta_{i,j}). \]
\[ \text{(B.7)} \]

Also, multiplying the two inequalities in (B.1) by \( (1 - \theta_{i,j}) \) and \( \theta_{i,j} \), respectively, and adding gives

\[ (\lambda_{i,j}(1 - \theta_{i,j}) + \mu_{i,j} \theta_{i,j}) \geq (\lambda_{i,j}^{u}(1 - \theta_{i,j}) + \mu_{i,j}^{u} \theta_{i,j}). \]
\[ \text{(B.8)} \]

Adding (B.7) and (B.8) gives (RHS of (B.6) > 0), hence from (B.6) \( \lambda_{i-1,j} - \lambda_{i,j}^{u} = \text{(RHS of (B.6)} \geq 0 \), which yields the first inequality in (B.2).

To obtain the second inequality in (B.2) (i.e., \( \mu_{i,j} > \mu_{i,j}^{u} \)) we note that on a utilized path \( d_{i} > 0 \), hence (15) becomes \( \lambda_{i,j} = \mu_{i,j} \). Also, by definition \( \lambda_{i,j}^{u} = \mu_{i,j}^{u} \). Subtracting gives \( (\lambda_{i,j} - \lambda_{i,j}^{u}) = (\mu_{i,j} - \mu_{i,j}^{u}) \) and using the first inequality in (B.1) (i.e., \( (\mu_{i,j} - \mu_{i,j}^{u}) \geq 0 \)), the second inequality in (B.2).

To prove (B.4) we note that \( \lambda_{T-i,j} = k_{T} \) and \( \lambda_{T-i-1,j} \leq k_{T} \). To prove (B.5) we note that \( \mu_{id} = \mu_{id}^{u} \equiv 0 \), which completes the proof (by induction) that \( \lambda_{i,j} > \lambda_{i,j}^{u} \).

Therefore, \( \lambda_{i,j}^{u} > 0 \), from (20) we have

\[ \lambda_{i,j}^{u} = k_{i} + \lambda_{i+1,j}(1 - \theta_{i+1,j}) + \mu_{i+1,j} \theta_{i+1,j}. \]

But \( k_{i}, \theta_{i+1,j}, \) and \( (1 - \theta_{i+1,j}) \) are all nonnegative, hence if \( \lambda_{i+1,j}^{u} \geq 0 \) and \( \mu_{i+1,j}^{u} \geq 0 \) we have \( \lambda_{i,j}^{u} \geq 0 \).

Proceeding in this way, by recursion from arc \( j \) in period \( t \), along all utilized paths to the destination and period \( T \), we obtain \( \lambda_{i,j}^{u} > 0 \) if \( \lambda_{i,j}^{u} > 0 \) for all \( j \) and \( \lambda_{i,j}^{u} \geq 0 \) for all \( t \). But for period \( T \), (20) reduces to \( \lambda_{T-i,j} = k_{C_{T}} \), which is positive, and for the destination arc \( n \), \( \lambda_{n,j} = 0 \). This completes the proof.

**Proposition 8.** Here \( \Delta \lambda_{i,j} \) can be \( \leq 0 \) or \( > 0 \) for any \( j \) and \( t \). This holds irrespective of whether \( \Delta \lambda_{i,j} \) and/or \( \Delta \lambda_{i,j}^{u} \) are negative, or positive, and irrespective of whether the volume on arc \( j \) is increasing \( (x_{i,j} > x_{i-1,j}) \) or decreasing \( (x_{i,j} < x_{i-1,j}) \).

**Proof.** To demonstrate this we need only give examples of \( \Delta \lambda_{i,j} < \Delta \lambda_{i,j}^{u} \) and \( \Delta \lambda_{i,j} > \Delta \lambda_{i,j}^{u} \), without making any assumption about the sign of \( \Delta \lambda_{i,j} \), \( \Delta \lambda_{i,j}^{u} \), or \( (x_{i,j} < x_{i-1,j}) \). We first state a case in which \( \Delta \lambda_{i,j} < \Delta \lambda_{i,j}^{u} \). Let arc \( j \) at time \( t \) be saturated, i.e., \( g_{i,j}^{u}(x_{i,j}) = 0 \). Then this reduces (17) to \( \lambda_{i-1,j} = (k_{C} + \lambda_{i,j}) \), hence

\[ (\lambda_{i,j} - \lambda_{i-1,j}) = \Delta \lambda_{i,j} = -k_{C}. \]

Rearranging (20) gives

\[ \Delta \lambda_{i-1,j} = (\lambda_{i,j} - \mu_{i,j} \theta_{i,j} - k_{C_{i}}. \]

and subtracting from \( (\Delta \lambda_{i,j} = -k_{C_{i}}) \) gives

\[ (\Delta \lambda_{i,j} - \Delta \lambda_{i,j}^{u}) = -(\lambda_{i,j} - \mu_{i,j} \theta_{i,j}) \]

Since \( (\lambda_{i,j} - \mu_{i,j} \theta_{i,j}) \) is normally positive and \( \theta_{i,j} \) is positive we have \( \Delta \lambda_{i,j} < \Delta \lambda_{i,j}^{u} \).

We now state a simple case in which \( \Delta \lambda_{i,j} > \Delta \lambda_{i,j}^{u} \).

Let \( j \) be an arc pointing into a destination node, in which case \( \mu_{i,j} \equiv 0 \) for all \( t \). Also let \( g_{i,j}^{u}(x_{i,j}) = b_{x_{i,j}} \), where \( b \) is a constant, up to the current value of \( x_{i,j} \). Then

\[ \theta_{i,j} = g_{i,j}(x_{i,j})/x_{i,j} = 1/g_{i,j}^{u}(x_{i,j}) \]

so that (20) reduces to

\[ (\lambda_{i,j} - \lambda_{i,j}^{u}) = \lambda_{i,j} = (-k_{C} + \theta_{i,j}) \]

and 17 reduces to \( \Delta \lambda_{i,j} = (-k_{C} + \theta_{i,j} \lambda_{i,j}) \) hence,

\[ (\Delta \lambda_{i,j} - \Delta \lambda_{i,j}^{u}) = \theta_{i,j}(\lambda_{i,j} - \lambda_{i,j}^{u}) \]

But \( \lambda_{i,j} > \lambda_{i,j}^{u} \) from Proposition 7, hence \( \Delta \lambda_{i,j} > \Delta \lambda_{i,j}^{u} \).

**REFERENCES**


