

Congestion Pricing under Operational, Supply-Side Uncertainty

Stephen Boyles Kara Kockelman S. Travis Waller

November 17, 2008

Abstract

This paper is concerned with finding first-best tolls in static transportation networks with day-to-day variation in network capacity, as accounted for by changes in the volume-delay function. The key question in addressing this problem is that of *information*, namely, which agents have access to what information when making decisions. In this work, travelers are assumed to be either fully informed about network conditions before embarking on travel, or having no information except the probability distributions; likewise, the network manager (toll-setter) is either able to vary tolls in response to realized network conditions, or must apply the same tolls every day. Further, travelers' preference for reliable travel is accounted for, representing risk aversion in the face of uncertainty. For each of the scenarios implied by combinations of these assumptions, we present methods to determine system-optimal link prices. A demonstration is provided, using the Sioux Falls test network, suggesting that attempts to incorporate uncertainty into nonresponsive tolls involve significantly higher prices.

1 Introduction

In recent years, pricing of highway driving (tolling) has attracted much political and institutional attention for a variety of reasons, including its potential as an alternate revenue stream, the introduction of technologies allowing efficient toll collection and dynamic pricing, and consideration of public-private partnerships. To support the process of determining appropriate prices, a large amount of research has been conducted to provide guidance on how users respond to prices, and how they should be set to achieve particular objectives. From the standpoint of maximizing social welfare, the fundamental notion, originated by Pigou (1920), is that economic efficiency occurs when the cost faced by each traveler equals the marginal social cost of his or her trip.

Traditionally, this marginal cost is determined by assuming a separable and differentiable volume-delay function (VDF) mapping travel demand to travel delay on each roadway segment, a homogeneous population of user-optimizing travelers with the same value of time, and commonly-known network structure

and travel demand. Within this framework, the marginal cost of tripmaking is readily calculated, along with the associated Pigouvian tolls.

One significant assumption in this process is the determinism of the transportation network. Uncertainty pervades both transportation planning and operations, and researchers are realizing the importance of explicitly incorporating this into transportation models. Uncertainty exists both in the short run, at the “operational” level (due to incidents, weather conditions, fluctuations in travel demand, etc.), and in the long run, “planning” level, due to imperfect prediction of future land use and economic conditions, and impacts parameters, both on the supply and demand sides of transportation. In this paper, the focus is on operational, supply-side uncertainty including, for instance, the impact of incidents, while excluding forecasting errors, or variations in travel demand.

This type of uncertainty directly leads to unreliability in travel times, which has a clear impact on user behavior. At the very least, the assumption that travelers seek to minimize travel time is ill-defined, since travel time is a random variable. The simplest extension is to assume that travelers seek to minimize *expected* travel time, but this is problematic because of an implicit assumption of risk neutrality. That is, under this assumption, travelers are indifferent between a trip which takes 30 minutes (with absolute certainty), and one which can either take 10 minutes or 50 minutes, with equal probability. For almost any conceivable trip purpose, the former is much preferred, since the consequences of late arrival are typically much worse than the benefits of early arrival. In fact, due to this asymmetry, it is reasonable to expect a traveler would even prefer a slightly longer trip, say, 35 minutes, if this travel time could be guaranteed.

However, this concept has not yet been integrated into pricing models intended for use in large-scale networks. Further, given recent technological advances, network operators and planners wonder whether prices should be dynamically varied in response to traffic incidents or other disruptions, and, if so, how this variation should occur. For instance, one might argue that tolls on a facility should increase if an incident occurs, to discourage additional vehicles from entering and exacerbating the resulting congestion. However, in response, one might argue that users paying a higher toll should expect a higher level of service, as this is one of the usual arguments provided to gain public support for congestion pricing. Wouldn't travelers resent paying a higher toll, while most likely still experiencing greater-than-average delay? Or is there some way to account for uncertainty on a daily basis without varying tolls?

As with most problems concerning uncertainty, the question of information is key: who knows what, when they make their decisions? For instance, the issue of resentment for higher tolls during an incident is greatly decreased if operators can communicate to motorists the presence and severity of the incident. This research presents four possible scenarios relating to the information available to motorists when choosing a travel route, and to the ability of the network manager to adjust the toll in response to network conditions.

The key contribution of this work is the development of pricing methods to apply in the presence of operational supply uncertainty and risk-averse travelers, for several information provision scenarios. The remainder of this paper

is organized as follows: Section 2 discusses prior literature related to pricing, travel time uncertainty, and user attitudes to risk. Section 3 describes the modeling approach, introducing appropriate notation and defining the four information scenarios. Section 4 presents solution methods for each of these scenarios. The models described thus far make a number of simplifications, and Section 5 discuss how they can be adapted to account for correlated link states, user heterogeneity, and elastic demand. Section 6 demonstrates the basic model using the well-known Sioux Falls test network, and suggests that constant tolls should not be used to address nonrecurring congestion. Finally, Section 7 concludes the paper, and summarizes the key findings.

2 Literature Review

This section summarizes prior related work, focusing on three areas: pricing under uncertain network conditions; the impact of reliability on route choice; and how pricing and reliability interact. In this light, the contribution of this paper should be more apparent, and is briefly discussed at the end of the section.

The question of how to appropriately price freeway facilities in uncertain environments is still very open. Yang (1999a) considered the problem of determining optimal prices when users behave according to the stochastic user equilibrium principle, where uncertainty lies in user perception, rather than system conditions. It is known that there need not exist a set of tolls that can drive a stochastic user equilibrium traffic flow pattern to a system optimal one (Akamatsu and Kuwahara, 1988; Smith et al., 1994). Yang (1999b) also considered how road pricing can be combined with advanced traveler information systems which inform users of system conditions. A number of numerical experiments were performed in a small test network, from which the author concluded that the two technologies “complement each other and that their joint implementation can reduce travel time more efficiently.” Separately, de Palma and Lindsey (1998) considered information provision under three different scenarios: free access, non-responsive congestion pricing, and dynamic pricing based on congestion levels. These authors explicitly considered capacity uncertainty in all of their models, but in a simplified setting without network effects and multiple origins and destinations. Under these assumptions, when pricing is dynamic and responsive to congestion, these authors showed that better information always improves welfare. A key result of Mohring and Harwitz (1962) is that marginal-cost pricing generates enough revenue to provide socially-optimal facility capacity; Lindsey (2008) showed that this result generalizes to the case of uncertain capacity if drivers are perfectly informed and tolls are responsive, or under imperfect information if tolls are set according to the same information drivers have, and if the price elasticity of demand does not vary with system conditions.

It is clear that reliability plays a significant role in route choice decisions; however, there is no consensus on how “reliability” should be defined. Usually, this is done in relation to the distribution of possible path costs. For instance,

Small et al. (2005) and Liu et al. (2007) used the difference between the 80th- and 50th-percentile travel times, while Pinjari and Bhat (2006) used the maximum additional time that could be needed, compared to a typical case. Gao (2005), on the other hand, assumed a piecewise-linear utility function to model risk aversion. de Palma and Picard (2005) considered four utility function specifications to represent risk aversion: penalizing the standard deviation of travel time, penalizing travel time variance, constant relative risk aversion, and constant absolute risk aversion. Bates et al. (2001) and Noland and Polak (2002) provided overviews of theoretical and empirical research in travelers' valuations of travel time reliability. Typically, travelers' sensitivity to reliability is comparable to their sensitivity to increased travel time; for instance, Small et al. (2005) estimated a \$21.46/hr value of time, and a \$19.56/hr value of reliability, using data from SR-91 in California.

In contrast to the utility-based methods above, Avineri and Prashker (2003) accounted for uncertainty in route choice using cumulative prospect theory, the Fudenberg-Levine learning model (Fudenberg and Levine, 1998), a behavioral "reinforcement learning" model, and a novel cumulative prospect theory learning model. Chan and Lam (2005) took a completely different approach, using a novel concept of user equilibrium based on "path preference indices."

Several researchers have studied the interaction between pricing and facility reliability. The research in this area has been descriptive (attempting to evaluate how pricing affects facility reliability, or studying pricing to discern how travelers value reliability) rather than prescriptive (how should prices be set to maximize reliability or traveler welfare). Supernak et al. (2003) performed a before-after study of the I-15 FasTrak value pricing project in San Diego, California, looking specifically at changes in travel time and travel reliability, measured as the 99th-percentile of travel time. Using this definition, they found substantial improvements in reliability after implementation. Liu et al. (2004) used freeway loop data from California State Route 91 to estimate a random-parameters logit model with two alternatives, free and tolled lanes. Travel time, reliability (defined as the difference between the 80th- and 50th-percentile travel times, approximately one standard deviation in several common probability distributions.), and toll amount are used as alternative-specific variables. They applied a genetic algorithm to estimate the logit parameters, resulting in an estimated value of time of \$13/hour and an estimated value of reliability of \$21/hour. Brownstone and Small (2005) also considered the I-15 and SR-91 project, and used both stated and revealed preference data. They estimated a relatively high value of time (between \$20/hour and \$40/hour for the morning commute) based on revealed preference data, and a much lower value (around \$12/hour) from stated preference surveys.

Although considerable research exists on pricing and network routing under uncertainty, relatively little research combines the two, especially regarding travelers' risk attitudes and/or valuation of reliable travel. Further, that which has been done has typically involved simplified settings and small networks, which admit important analytical results but which are less useful in guiding implementation of pricing policies. Emmerink et al. (1996) showed that no subsidy

towards information provision is needed to maximize social welfare, given first-best congestion pricing and costly information provision. Verhoef et al. (1996) simulated a two-link network under different pricing and information scenarios, concluding that information provision and “flat” (unresponsive) tolls are nearly as effective as perfectly responsive tolls. Kobayashi and Do (2005) considered a simple network with non-overlapping routes and a single origin-destination pair, and showed that perfect information and *ex post* tolls maximize social welfare.

In this light, this paper’s primary contribution is the development of models to identify optimal tolls in large-scale networks, when roadway capacity is stochastic. In these models, corresponding to different information scenarios, route choice is endogenous (i.e., traffic assignment and equilibrium are included), and prices inducing system-optimal (or approximately system-optimal) link flows are sought.

3 Problem Statement

This section mathematically describes the network and pricing model assumptions, along with four information scenarios. Section 3.1 defines the basic problem and introduces notation that will be used throughout this document. Section 3.2 addresses the issue of how to model users’ valuation of reliability in route choice, and reviews the equilibrium concepts that will be applied to determine user response to a set of tolls. Section 3.3 describes the network manager’s goal, and section 3.4 defines the information scenarios we analyze. Three strong assumptions — independence of link states, inelastic demand, and user homogeneity — are taken in this section to emphasize the key points of the basic model; the implications of relaxing these are discussed in Section 5.

3.1 Notation

Generally, pricing problems are a type of Stackelberg game, in which a regulator acts as a “leader” by establishing a set of tolls, to which individual drivers (“followers”) respond by choosing preferred routes. We adopt the same perspective, but with the additional complication of uncertain network conditions. As described in Section 3.2, a generalized cost function is assumed for travelers, accounting for average travel time, reliability, and monetary tolls. The goal of the network manager is to choose tolls so as to bring the user equilibrium and system optimal link flows into alignment (i.e., incorporating externalities into individual costs).

Specifically, consider a network $G = (N, A, D)$ consisting of a set of nodes N , a set of arcs A , and a set of origin-destination (OD) demand pairs D (OD pair, for short). Each arc $(i, j) \in A$ can exist in a variety of *states* $S_{ij} = \{1, 2, \dots, s_{ij}\}$. For each state s of arc (i, j) , a positive and increasing travel time function $t_{ij}^s(x_{ij})$ relates the demand x_{ij} for arc (i, j) to the experienced travel time in state s (for instance, one state might represent a slightly lower capacity due to wet roadway conditions, while another represents a significantly lower capacity due to the

temporary loss of a lane). State s is presumed to occur with probability $p(s_{ij})$, independent of the state of any other arc. This induces a probability space (Ω, \Pr) where Ω consists of all realizations of the *network* (i.e., each $\omega \in \Omega$ is an element of $\times_{(i,j) \in A} S_{ij}$), and $\Pr(\omega) = \prod_{(i,j) \in A} p(s_{ij}^\omega)$.

Note that the independence assumption limits the models' ability to represent, for instance, secondary incidents which occur as a result of an earlier incident, or weather events which simultaneously affect multiple links. In Section 5, the implications of relaxing this assumption are discussed; briefly, relaxing this assumption only introduces difficulties if users do not receive travel information before departing. In the interim, the independence assumption is adopted to simplify the exposition.

Other notation is defined as follows: the decision variables τ_{ij} represent the monetary toll on link (i, j) . A fixed and known amount of demand d_{rs} , assumed to be infinitely divisible, is associated with each OD pair $(r, s) \in D$. Let Π_{rs} be the set of simple paths connecting origin r to destination s , $\Pi = \cup_{(r,s) \in D} \Pi_{rs}$, and δ_{ij}^π be an indicator variable set to unity if arc $(i, j) \in \pi$ for all $\pi \in \Pi$, and zero otherwise.

Depending on the information provision scenario, the tolls and link flows may vary according to the network realization ω ; when needed, a superscript will denote which realization a toll or flow value corresponds to.

3.2 User Behavior

All travelers are assumed to be homogeneous, and value travel on any path π according to an additive generalized cost function C , so the cost associated with any path $C^\pi = \sum_{(i,j) \in A} \delta_{ij}^\pi c_{ij}$, where c_{ij} is the cost associated with travel on any link (i, j) .

When travelers know link states before departing, the link cost function is simply

$$c_{ij} = VOTT \times t_{ij} + \tau_{ij} \quad (1)$$

where $VOTT$ represents the value of travel time. On the other hand, if link states are unknown to travelers when they choose routes, the cost function is

$$c_{ij} = VOATT \times E[\tilde{t}_{ij}] + VOTR \times Var[\tilde{t}_{ij}] + \tau_{ij} \quad (2)$$

where $VOATT$ and $VOTR$ represent the value of average travel time (not necessarily the same as $VOTT$), and the value of travel reliability, respectively. In either case, users are assumed to correctly perceive travel times, tolls, and, in the latter case, the travel time distributions as well. That is, there is no need to adopt a stochastic user equilibrium framework, because the model assumes users' attitudes towards uncertain conditions are exactly as described by these utility functions. Note that in the case of information provision, cost function (1) is a special case of (2), in which the variance in link cost vanishes. Thus, these two utility functions are consistent, and meaningful comparisons can be made between tolls derived for both scenarios.

This specification only considers within-day travel time uncertainty: even if travel times vary widely from day to day, equation 1 is used as long as travelers learn the exact realization before departing. That is to say, the inherent value of stable travel times over multiple days, such as the establishment of routine habits, is excluded from consideration. Rather, the focus in this paper is the cost of imperfect knowledge of travel times on a given day, leading to earlier departure times (leaving a “safety margin”) or running the risk of late arrival.

As described in Section 2, some researchers prefer to use standard deviation instead of variance, since it has common units with expected travel time. In this paper, we opt to use variance for three reasons: first, mean-variance models are commonly used to model risk in domains such as finance (see, for instance, Markowitz, 1952; Sternbach, 2001); second, it is more convenient mathematically, since variances add linearly under the independence assumption (i.e., $Var[A + B] = Var[A] + Var[B]$), allowing ready computation of path variance; finally, since variance is the square of standard deviation, this model places increasingly greater weight on travel reliability as travel times become more uncertain.

All travelers are assumed to be self-optimizers who independently choose routes to minimize their own generalized travel cost; in a deterministic and toll-free environment, this is identical to Wardrop’s first condition and the classic deterministic user equilibrium problem. It should be emphasized that our model describes a long-term (planning-level) equilibrium, despite consideration of day-to-day uncertainty. The key notion is that day-to-day uncertainty impacts habitual decisions which long-term models seek to capture. The choice of a long-term framework justifies the assumptions of users’ knowledge of the network, gained from experience.

As a side note, one may not need to explicitly sum over all arc states when evaluating $E[\tilde{t}_{ij}]$ and $Var[\tilde{t}_{ij}]$, depending on the cost function. For instance, assuming a standard Bureau of Public Roads cost function of the form $\tilde{t}_{ij} = t_{ij}^0(1 + \alpha(x_{ij}/\tilde{c}_{ij})^\beta)$ in which only the capacity parameter \tilde{c}_{ij} is stochastic, it is readily verified that $E[\tilde{t}_{ij}] = t_{ij}^0(1 + \alpha\phi_{ij}x_{ij}^\beta)$ and $Var[\tilde{t}_{ij}] = \theta_{ij}(\alpha t_{ij}^0 x_{ij}^\beta)^2$, where $\phi_{ij} = \sum_{s \in S_{ij}} c_{ij}^{-\beta} p(s_{ij})$ and $\theta_{ij} = \sum_{s \in S_{ij}} c_{ij}^{-2\beta} p(s_{ij}) - \phi_{ij}^2$ are link-specific constants that need only be calculated once, independent of demand values and route choices.

3.3 Network Manager Behavior

In our model, the network manager’s goal is to maximize the total *travel time*-related costs experienced by travelers; that is, when users learn link states before departing, the network manager seeks to minimize

$$\sum_{(i,j) \in A} x_{ij} (VOTT \times E[\tilde{t}_{ij}]) \tag{3}$$

In contrast, when users are ignorant of link states when choosing routes, the network manager minimizes

$$\sum_{(i,j) \in A} x_{ij} (VOATT \times E[\tilde{t}_{ij}] + VOTR \times Var[\tilde{t}_{ij}]) \quad (4)$$

This is done by setting the tolls τ in such a manner as to bring the user and system objectives into alignment. Note that the network manager’s goal does not include minimizing the toll-related costs. This assumes that toll revenues are effectively returned to travelers with minimal administrative burden, perhaps through additional infrastructure spending or reduced taxation.

3.4 Information Scenarios

The question of which agents have access to what information plays a defining role in determining the structure and results in a stochastic optimization model. In this problem, there are two types of agent: the network manager, who establishes the tolls, and the users, who choose routes.

Initially, we consider two information scenarios for each agent (leading to four scenarios in total): a “no information” case, in which the agent is unaware of the network realization before making the decision, and a “fully informed” case, in which the agent learns the exact network realization. This information is assumed to be perfectly accurate and fully trusted by all agents.

For the manager, the “no information” case is identical to one in which tolls cannot vary from day to day; for this reason, the manager’s information is denoted as either RT (responsive tolls, for full information) or UT (unresponsive tolls, for no information or when responsive tolling is impossible). For the users, the “no information” case implies that the route must be chosen before learning the network realization and the tolls; the “full information” case implies that the tolls and network realization are both known. These user information scenarios are denoted NI and FI, respectively. Thus, the four information scenarios here are FI/UT, FI/RT, NI/UT, and NI/RT (indicating the users’ information first and the regulator’s information second):

Fully Informed/Unresponsive Tolls This scenario represents a case in which users learn the network realization before choosing routes, but tolls cannot vary in response to their choices or the network realization. This can occur either because the regulator is unaware of network conditions, or because the regulator is not allowed to change the tolls. More precisely, users learn the prevailing travel time functions and their relevant parameters (such as capacity), and the equilibrium state arising from this common knowledge is sought. Thus, the flows \mathbf{x}^ω vary according to the network realization, but the tolls τ do not.

Fully Informed/Responsive Tolls This scenario represents maximum information for all decision makers in the problem: travelers learn conditions

throughout the network before choosing a route, as does the network manager before setting tolls. This is similar to the FI/UT case, except that the tolls τ are also allowed to vary according to the network realization.

No Information/Unresponsive Tolls This information scenario represents one in which neither users nor the manager can vary their decisions according to the network realization. In this case, although the tolls are known and fixed from day to day, users are unaware of the network realization when making the routing decision, and thus have no reason to vary their decision from day to day.

No Information/Responsive Tolls This scenario represents the case in which users are unaware of the network realization, even though the network manager can vary the tolls responsively. However, varying tolls cannot provide any additional benefit to users if they do not learn of them before they choose a route. Essentially, this scenario is identical to NI/UT, since the manager cannot induce a superior flow pattern by varying tolls if users cannot respond in turn. More formally, if travelers are unaware of tolls when choosing a route, equation (2) must be modified, e.g., $u_{ij} = -VOATT \times E[\tilde{t}_{ij}] - VOTR \times Var[\tilde{t}_{ij}] - f(\tilde{\tau}_{ij}, \Omega, Pr)$ for some function f (including, for instance, the expected value and variance of the tolls $\tilde{\tau}$ which are now perceived as random variables), and let $\mathbf{x}(\tau)$ represent the equilibrium link flow vector obtained for tolls τ over all network realizations. The same flow vector can be replicated under the NI/UT scenario by defining a random variable \tilde{v} which takes the values τ_{ij}^{ω} with probabilities $Pr(\omega)$, and setting tolls $\tau_{ij} = f(v, \Omega, Pr)$. The generalized cost on each link is the same under this construction as in the NI/RT scenario, and thus \mathbf{x} remains an equilibrium.

Another reasonable information scenario is the case in which users learn the state and toll of an arc only upon reaching its upstream node, and are allowed to re-route depending on what they learn. While algorithms exist to model this behavior at the individual level (Polychronopoulos and Tsitsiklis, 1996; Waller and Ziliaskopoulos, 2002; Gao, 2005), an efficient algorithm to find a multiuser equilibrium (let alone optimal tolls) under this behavioral assumption is beyond the current state of the art and, therefore, not pursued here.

Note that in all of these scenarios, all agents are assumed to know the travel demand, and the probability distributions governing all link states. This is the direct analogue of the full information assumption implicit in most equilibrium models.

4 Solution Methods

This section describes methods for finding tolls that bring the user and system objectives into alignment, for the three information scenarios FI/UT, FI/RT, and NI/UT. (As demonstrated in the previous section, NI/RT is a special case

of NI/UT, and need not be considered separately). All of these are based on the Pigouvian principle that externalities should be incorporated into user costs or, equivalently, that average cost equal the marginal social cost. Since the marginal cost of travel on a link (i, j) is $d(x_{ij}c_{ij})/dx_{ij} = c_{ij} + x_{ij}(dc_{ij}/dx_{ij})$, the Pigouvian toll is $x_{ij}(dc_{ij}/dx_{ij})$ with the cost function appropriate to the information scenario, and system optimal flows \mathbf{x} .

4.1 Fully Informed/Unresponsive Tolls

The FI/UT scenario is the most complicated to solve, for several reasons. First, the constraint that the tolls must be the same for all network realizations prevents a simple decomposition by network realization. Second, the requirement that the flows for every network realization be in equilibrium imposes non-convexity on the toll-setting problem (see, for instance, Labbé et al., 1998), a problem confounded by the nonlinearity of travel time functions, which makes it unlikely that a globally optimal solution can be found. Third, since the flows vary according to the network realization, and since the number of network realizations is very large, computation of the objective function for even a single set of tolls is nontrivial. Essentially, the desired toll vector solves the program

$$\min_{\boldsymbol{\tau}} \sum_{\omega \in \Omega} \Pr(\Omega) \sum_{(i,j) \in A} t_{ij}^{\omega}(x_{ij}^{\omega}) x_{ij}^{\omega} \quad (5)$$

$$\text{s.t. } \tau_{ij} \geq 0 \quad \forall (i, j) \in A \quad (6)$$

$$\mathbf{x}^{\omega} \in Eq(VOTT \times \mathbf{t}^{\omega} + \boldsymbol{\tau}) \quad \forall \omega \in \Omega \quad (7)$$

where $Eq(VOTT \times \mathbf{t}^{\omega} + \boldsymbol{\tau})$ represents the set of user equilibrium link flows for delay functions \mathbf{t}^{ω} and tolls $\boldsymbol{\tau}$. This is a nonlinear mathematical program with equilibrium constraints (MPEC), which is known to be difficult to solve, for the reasons mentioned above. For this reason, approximately optimal tolls are sought. One option is to use a generic metaheuristic, such as simulated annealing or tabu search. Another choice is to use problem-specific heuristics, two of which are described below.

Heuristic 1 (H1) is to use simple averaging: for each scenario ω , first-best marginal-cost tolls are found, and the final toll vector is obtained by taking an average of these, weighted according to the probability of each scenario occurring. If the number of scenarios is large, a selection of these can be obtained using Monte Carlo simulation or another sampling technique. In addition to its simplicity, this heuristic is also motivated by a result in Lindsey (2008) stating that, for a single facility, the optimal unresponsive toll is the average of responsive tolls; thus, it is plausible that such tolls may be near-optimal in a network setting as well.

Heuristic 2 (H2) is somewhat more involved, and allows tolls to vary by realization, penalizing this variation in the objective function with a positive

constant M :

$$\min_{\boldsymbol{\tau}} \sum_{\omega \in \Omega} \Pr(\Omega) \sum_{(i,j) \in A} [t_{ij}^{\omega}(x_{ij}^{\omega})x_{ij}^{\omega} + M(\tau_{ij}^{\omega} - \bar{\tau}_{ij})^2] \quad (8)$$

$$\text{s.t. } \tau_{ij}^{\omega} \geq 0 \quad \forall (i,j) \in A, \omega \in \Omega \quad (9)$$

$$\mathbf{x}^{\omega} \in Eq(VOTT \times \mathbf{t}^{\omega} + \boldsymbol{\tau}^{\omega}) \quad \forall \omega \in \Omega \quad (10)$$

where $\bar{\tau}_{ij}$ is the average toll on arc (i,j) across all realizations. The problem is then decomposed by realization, and solved approximately by linearizing the objective function. The realization-specific subproblems are

$$\min_{\boldsymbol{\tau}^{\omega}} [t_{ij}^{\omega}(x_{ij}^{\omega})x_{ij}^{\omega} + M(\tau_{ij}^{\omega} - \bar{\tau}_{ij})^2] \quad (11)$$

$$\text{s.t. } \tau_{ij}^{\omega} \geq 0 \quad \forall (i,j) \in A \quad (12)$$

$$\mathbf{x}^{\omega} \in Eq(VOTT \times \mathbf{t}^{\omega} + \boldsymbol{\tau}^{\omega}) \quad (13)$$

which are solved if

$$\frac{d \left(\sum_{(k,\ell) \in A} x_{k\ell}^{\omega} t_{k\ell}^{\omega}(x_{k\ell}^{\omega}) \right)}{d\tau_{ij}^{\omega}} = -2M(\tau_{ij}^{\omega} - \bar{\tau}_{ij}) \quad (14)$$

for all $(i,j) \in A$. Applying the chain rule to the left-hand side, an analytical solution for each τ_{ij} is obtained:

$$\tau_{ij}^{\omega} = \bar{\tau}_{ij} - \frac{1}{2M} \sum_{(k,\ell) \in A} \frac{dx_{k\ell}}{d\tau_{ij}} \left[t_{k\ell}^{\omega}(x_{k\ell}^{\omega}) + x_{k\ell}^{\omega} \frac{dt_{k\ell}^{\omega}}{dx_{k\ell}^{\omega}} \right] \quad (15)$$

where the $dx_{ij}/d\tau_{kl}$'s are estimated by perturbing the toll on each arc slightly and observing the resulting change in equilibrium link flows. This suggests an iterative procedure in which subproblems are successively solved, with increasing values of the penalty constant M (clearly, as $M \rightarrow \infty$ the realization-specific tolls converge to a common value, which is returned as the solution). This is shown formally in Algorithm 1, where M_0 and ϵ respectively denote the initial value of the penalty constant, and the convergence criterion. If $|\Omega|$ is large, a sample of realizations can be used in lieu of the entire set.

4.2 Fully Informed/Responsive Tolls

In the FI/RT scenario, tolls are allowed to vary for different network realizations, enabling a natural decomposition of the problem: since the variance terms in the cost function vanish, it suffices to find the system optimal link flows using the travel time functions from each network realization ω , and the optimal tolls that they imply.

Although this leads to optimal tolls for every realization, enumerating every realization is prohibitively difficult for networks of any reasonable size. From a practical standpoint, this may not be a significant issue since tolls can be

Algorithm 1 Heuristic 2 for FI/UT

```
1: {Initialization}
2: for all  $\omega \in \Omega$  do
3:    $\tau^\omega \leftarrow$  first-best marginal cost tolls for scenario  $\omega$ .
4: end for
5:  $\bar{\tau} \leftarrow \sum_{\omega \in \Omega} \Pr(\omega) \tau^\omega$ 
6:  $M \leftarrow M_0$ 
7: {Iteration}
8: while  $\max_{(i,j), \omega} |\tau_{ij}^\omega - \bar{\tau}_{ij}| > \epsilon$  do
9:   for all  $\omega \in \Omega, (i, j) \in A$  do
10:     $\tau_{ij} \leftarrow \bar{\tau}_{ij} - \frac{1}{2M} \sum_{(k,\ell) \in A} \frac{dx_{k\ell}}{d\tau_{ij}} \left[ t_{k\ell}^\omega(x_{k\ell}^\omega) + x_{k\ell}^\omega \frac{dt_{k\ell}^\omega}{dx_{k\ell}^\omega} \right]$ 
11:   end for
12:    $M \leftarrow 2M$ 
13: end while
14: return  $\bar{\tau}$ 
```

determined as needed (rather than all calculated beforehand and stored) and/or optimal toll policies can be determined for the most likely scenarios.

Nevertheless, from the standpoint of research and planning, it is desirable to make statements about the overall costs to travelers under this information provision scenario. Although variance disappears when solving the individual subproblems, it may be useful to calculate an *a priori* variance, representing the uncertainty that exists prior to learning the network states on a given day. To this end, a sampling procedure can be used to approximate these quantities.

4.3 No Information/Unresponsive Tolls

The NI/UT scenario is the simplest case, since only one vector of network flows and one vector of tolls is needed. Solving for system-optimal link flows and marginal-cost prices using cost function (2) gives optimal tolls for this information scenario.

5 Extensions to the Basic Model

The previous section describes pricing models in a simplified setting to allow the main points to be made more clearly, and to show that information scenarios such as NI/UT are inherently difficult to solve even without sophisticated traffic or behavioral models. Still, this basic framework can be enhanced in several directions, and this section discusses three such extensions: correlated link states, user heterogeneity, and elastic demand.

5.1 Correlated Link States

In preceding sections, it is assumed that link states are independent in order to facilitate computation of path travel time variance. However, this clearly limits the model's ability to represent phenomena such as weather, special events, or even incidents (which create queues which affect multiple links). However, the independence assumption can be relaxed.

For the FI cases, users' utility function only depends on path travel times, obviating the need to calculate variances. Still, in large networks where it is impractical to enumerate all network states, a more sophisticated sampling scheme may be required to account for the dependence between link states. While the well-known inverse transform sampling procedure (see, for instance, Casella and Berger, 2002), applied separately to each link, suffices for the independent case, this method will not simulate dependence. Falk (1999) proposes an approximate method for generating vectors of uniform random variables with an arbitrary covariance matrix which should be suitable for this application.

A more serious complication arises for the NI case, in that travel costs are no longer additive over links. One expects that the cost of a *path* π should be

$$C^\pi = VOATT \times E[\tilde{t}^\pi] + VOTR \times Var[\tilde{t}^\pi] + \tau^\pi \quad (16)$$

where \tilde{t}^π and τ^π are the (random) travel time and (deterministic) toll experienced on path π . This can be written in terms of link travel times and tolls as

$$C^\pi = \sum_{(i,j) \in A} \delta_{ij}^\pi (VOATT \times E[\tilde{t}_{ij}] + VOTR \times Var[\tilde{t}_{ij}] + \tau_{ij}) + \\ + VOTR \times \sum_{(i,j) \in A} \sum_{(k,l) \in A \setminus (i,j)} \delta_{ij}^\pi \delta_{kl}^\pi Cov[\tilde{t}_{ij}, \tilde{t}_{kl}] \quad (17)$$

where the latter term cannot be separated by arc. In the independent case, of course, $Cov[\tilde{t}_{ij}, \tilde{t}_{kl}] = 0$ for $(i,j) \neq (k,l)$ and this term vanishes, allowing the separation in (2).

This is problematic because the network manager's objective (4) must now explicitly account for every user's path choice, instead of simply their link choices. As the number of paths in realistic-sized networks is much larger than the number of links, and as user equilibrium solutions are not unique in path choices, minimizing the sum of user generalized cost becomes difficult. A reasonable and tractable alternative is to minimize the mean and variance of total system travel time $TSTT = \sum_{(i,j) \in A} x_{ij} \tilde{t}_{ij}$, weighted according to the values of travel time and travel reliability:

$$VOATT \times E[TSTT] + VOTR' \times Var[TSTT] = \\ = \sum_{(i,j) \in A} VOATT \times x_{ij} E[\tilde{t}_{ij}] + VOTR' \times x_{ij}^2 Var[\tilde{t}_{ij}] + \\ + \sum_{(i,j) \in A} \sum_{(k,l) \in A \setminus (i,j)} VOTR' \times x_{ij} x_{kl} Cov[\tilde{t}_{ij}, \tilde{t}_{kl}] \quad (18)$$

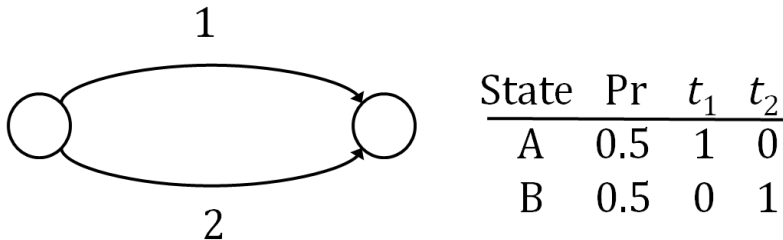


Figure 1: Variance in $TSTT$ is not equal to variance in individual travel times.

where $VOTR'$ is the network manager's value of reliability in $TSTT$. $VOTR'$ has different units than $VOTR$, indicating that this objective is not exactly the same as minimizing the sum of users' travel time and reliability costs. This point is reinforced in the network and flow pattern shown in Figure 1, where the two arcs' costs are perfectly correlated, so that if the cost of one is 1, the cost of the other is 0. If each link is used by one user, each user experiences a positive travel time variance, although the variance in $TSTT$ is zero. While this example is somewhat contrived, it shows that (18) is only an approximation of (4).

The system-optimal link flows \mathbf{x}^* with respect to this new objective is unique and easily found, since the covariance matrix is positive definite, implying that the objective (18) is convex. These link flows can then be decomposed into a vector of path flows \mathbf{h}^* ; the remaining question is how to set link tolls so that users have no incentive to deviate from these path flows. (In general, there is more than one path flow vector \mathbf{h}^* corresponding to link flows \mathbf{x}^* ; however, all such path flow vectors yield the system-optimal link flows and can be considered equivalent from this perspective.)

For each OD pair $(r, s) \in D$, the set of paths $\Pi_{r,s}$ is thus partitioned into two sets: $\hat{\Pi}_{r,s}$, corresponding to paths with a cost no greater than the highest-cost used path (that is, $\pi \in \hat{\Pi}_{r,s}$ if and only if there exists a path $\rho \in \Pi_{r,s}$ such that $h^\rho > 0$ and $C^\rho \geq C^\pi$), and $\bar{\Pi}_{r,s}$, corresponding to all other paths. Intuitively, one can interpret $\hat{\Pi}_{r,s}$ as the set of “competitive” paths, and $\bar{\Pi}_{r,s}$ as the set of “non-competitive” paths. Thus, we seek a vector of tolls $\boldsymbol{\tau}$ and a vector of OD costs $\boldsymbol{\kappa}$ satisfying the following system of linear inequalities:

$$C^\pi + \sum_{(i,j) \in A} \delta_{ij}^\pi \tau^\pi \geq \kappa_{rs} \quad \forall (r, s) \in D, \pi \in \bar{\Pi}_{r,s} \quad (19)$$

$$C^\pi + \sum_{(i,j) \in A} \delta_{ij}^\pi \tau^\pi = \kappa_{rs} \quad \forall (r, s) \in D, \pi \in \hat{\Pi}_{r,s} \quad (20)$$

$$\tau_{ij} \geq 0 \quad \forall (i, j) \in A \quad (21)$$

To show that a solution to equations always exists, we introduce an artificial objective function $\min_{\boldsymbol{\tau}, \boldsymbol{\kappa}} \mathbf{0} \cdot \boldsymbol{\tau} + \mathbf{0} \cdot \boldsymbol{\kappa}$ to create a linear program with con-

straints (19)–(21). The corresponding dual problem is

$$\min_{\lambda, \mu} \sum_{(r,s) \in D} \left(\sum_{\pi \in \bar{\Pi}_{r,s}} C^\pi \lambda^\pi + \sum_{\pi \in \bar{\Pi}_{r,s}} C^\pi \mu^\pi \right) \quad (22)$$

$$\text{s.t.} \quad \sum_{\pi \in \bar{\Pi}} \delta_{ij}^\pi \lambda^\pi + \sum_{\pi \in \bar{\Pi}} \delta_{ij}^\pi \mu^\pi \leq 0 \quad \forall (i, j) \in A \quad (23)$$

$$\sum_{\pi \in \bar{\Pi}_{r,s}} \lambda^\pi + \sum_{\pi \in \bar{\Pi}_{r,s}} \mu^\pi = 0 \quad \forall (r, s) \in D \quad (24)$$

$$\lambda^\pi \geq 0 \quad \forall \pi \in \bar{\Pi} \quad (25)$$

with multipliers λ and μ corresponding to constraints (19) and (20), respectively. Note that the constraints (24) imply $\sum_{\pi \in \bar{\Pi}_{r,s}} \lambda^\pi = -\sum_{\pi \in \bar{\Pi}_{r,s}} \mu^\pi$. Since the cost of every path in $\bar{\Pi}$ is strictly greater than the cost of every path in $\hat{\Pi}$, and since $\lambda^\pi \geq 0$ for all π , we have $\sum_{\pi \in \bar{\Pi}_{r,s}} \lambda^\pi + \sum_{\pi \in \bar{\Pi}_{r,s}} \mu^\pi \geq 0$ for all OD pairs (r, s) , implying that the objective function (22) is nonnegative and that the dual problem is not unbounded. Furthermore, the dual problem is feasible: $\mu = \mathbf{0}$ and $\lambda = \mathbf{0}$ satisfy all of the dual constraints. (In fact, this solution solves the dual problem optimally, but this is irrelevant because the primal objective function was chosen arbitrarily.) The existence of a feasible solution to the primal thus follows immediately from strong duality. Solving the corresponding Phase I linear program will produce a toll vector enforcing the desired flow pattern (for more details, see Bertsimas and Tsitsiklis, 1997); one may also choose to introduce another objective function at this point, and identify a toll vector optimizing a criterion such as total toll burden.

Note that this method, as stated, requires path enumeration. In practice, one can often restrict attention to a set of “reasonable” paths, such as those satisfying the efficiency condition suggested by Dial (1971), and define the set $\bar{\Pi}_{r,s}$ accordingly. Also, in the interest of keeping the toll vectors as small as possible, the link flow decomposition should ideally utilize low-cost paths. This is not trivial — when path costs are not an additive function of link costs, even finding the shortest path involves solving an integer program (Gabriel and Bernstein, 2000) — and is an interesting topic worthy of further study.

Still, the method presented in this section provides insight on how tolls should be set to account for disruptive events affecting multiple links, such as severe incidents or stormy weather.

5.2 User Heterogeneity

Users are not uniform in their valuation of travel time and reliability. Instead, one can imagine that the parameters $VOTT$, $VOATT$, and $VOTR$ can be represented as (possibly correlated) distributions over the population. The idea that these parameters vary in the population is both intuitive, and has been empirically demonstrated in multiple stated and revealed preference surveys. If the value of reliability is not considered (or if $VOTR$ is assumed to be an affine

function of $VOATT$), the bicriterion equilibrium and pricing framework presented in Dial (1996, 1997, 1999a,b) suffices for identifying welfare-maximizing tolls. For the “full information” cases, where path travel time variance vanishes, nothing further is needed.

On the other hand, including reliability as a third criterion (alongside toll and average travel time, as in the “no information” case) introduces a few more complications. With a few suitable modifications, Dial’s approach can be applied to the tricriterion case, as described in this section. The basic concept involves identifying a set of efficient paths and assigning trips accordingly, because only a small set of paths will be used by travelers regardless of their values of travel time and reliability. This assignment process is then applied iteratively to find a user equilibrium and welfare-maximizing tolls.

These assignments are performed using prevailing path attributes. That is, link tolls, travel time means, and travel time variances are temporarily assumed to be fixed and independent of traffic flow, and the efficient paths are those which are least-cost paths with respect to some values of $VOATT$ and $VOTR$. Each path π has an associated *tricriterion vector* \mathbf{P}^π , whose three components are the toll, mean travel time, and travel time variance on path π . Plotting all vectors \mathbf{P} in the first octant, the efficient paths are seen to be the lower extreme points of their convex hull.

One can identify efficient paths using several techniques from multiobjective optimization, such as weighted objective functions (Geoffrion, 1968), the ϵ -constraining method (Haimes et al., 1971), and a decomposition method using the Chebyshev metric (Eswaran et al., 1989). Since our utility functions are simply linear functions of the path attributes, the weighted objective function method as adapted by Dial (1996) generates these paths efficiently for the bicriterion problem.

Dial (1996) speculates that the efficient frontier of the tricriterion problem is a “triangulated convex surface, with each vertex representing a path.” This point requires careful definition, since the convex hull of the set of efficient paths may include faces adjoining more than three vertices. Consider a network consisting of four paths, whose tolls, mean travel times, and travel time variances are shown in Table 1. All of these paths are efficient, and their convex hull is a plane segment with four vertices, as seen in Figure 2. Of course, one can express this quadrilateral plane segment as the union of triangular plane segments (say, triangles P1-P2-P4 and P1-P3-P4), in which case the efficient frontier is trivially seen to consist of a convex union of triangles. This representation is useful algorithmically, and adopted throughout this section.

Once a set of efficient paths is identified, one must assign heterogeneous users to these paths. In the bicriterion case, travelers are partitioned according to their values of travel time, and Dial (1996) shows that each path can be associated with an interval in \mathbb{R}_+ , and each user chooses the path associated with the interval containing their own value of travel time. In the tricriterion case, each path is associated with a *region* in \mathbb{R}_+^2 ; each user chooses the path associated with the region containing their values of $VOATT$ and $VOTR$. Figure 3 shows these regions for the paths in Table 1.

Path	τ	$E[t]$	$Var[t]$
P1	2	0	1
P2	0	0	2
P3	2	1	0
P4	0	2	0

Table 1: Four paths used in tricriterion demonstration

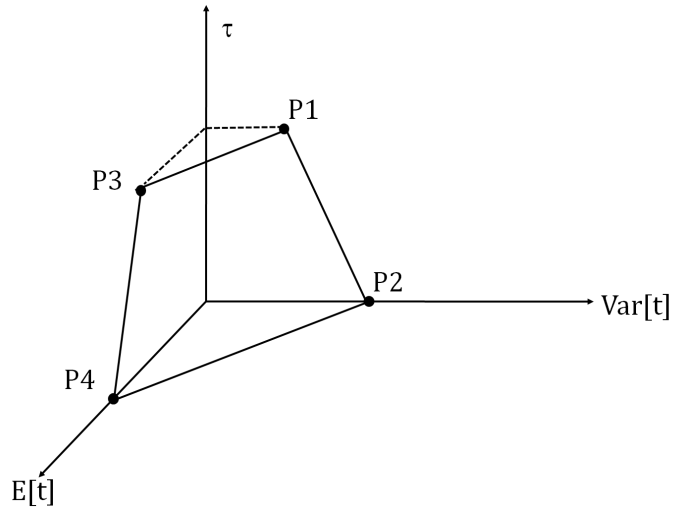


Figure 2: Four efficient paths and their convex hull

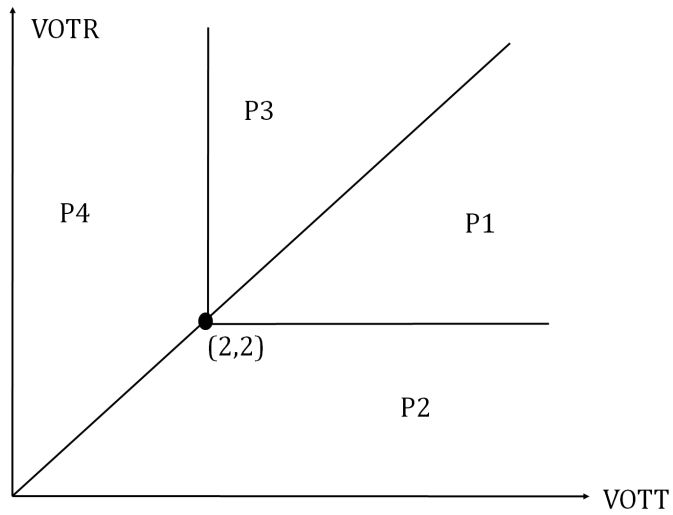


Figure 3: Regions of $VOTT-VOTR$ space corresponding to each efficient path (the “dual graph”)

Note that the vertices in Figure 2 correspond to regions in Figure 3, and vice versa; thus, these two graphs can be considered dual to each other. For instance, the point $(2, 2)$ in Figure 3 is adjacent to the regions corresponding to all four paths, indicating that when $VOATT = VOTR = 2$, these paths have equal costs. The corresponding region in Figure 2 is the plane segment P1-P2-P4-P3. One can verify that the vector $[\tau, E[t], Var[t]] = [2, 4, 4]$ is normal to this plane segment, suggesting that the plane segment represents $(VOTT, VOTR) = (4/2, 4/2) = (2, 2)$, thus demonstrating the correspondence with the dual graph.

A general procedure for identifying efficient paths can now be described. Given a triangular plane segment with vertices $\{\mathbf{P}^1, \mathbf{P}^2, \mathbf{P}^3\}$, the corresponding values of $VOATT$ and $VOTR$ can be obtained from a vector normal to the segment (a normal vector can easily be found by taking the cross product of $\mathbf{P}^2 - \mathbf{P}^1$ and $\mathbf{P}^3 - \mathbf{P}^1$). The least-cost path π for these particular values of $VOATT$ and $VOTR$ are identified, along with the corresponding criterion vector \mathbf{P}^π ; if this path is not already part of the efficient set, add it, and divide the plane segment into three new segments: $\{\mathbf{P}^1, \mathbf{P}^2, \mathbf{P}^\pi\}$, $\{\mathbf{P}^1, \mathbf{P}^3, \mathbf{P}^\pi\}$, and $\{\mathbf{P}^2, \mathbf{P}^3, \mathbf{P}^\pi\}$. These new plane segments are then recursively examined in the same way, until all efficient paths have been identified. It is worthwhile to update the dual graph at the same time, by eliminating the point corresponding to the original plane segment, adding the three points corresponding to the new plane segments, and updating the edges so that adjacent plane segments in the primal are directly connected in the dual.

One must be careful in initializing this algorithm. For the bicriterion case, it suffices to identify the least-cost and least-time paths, and use the slope of the line connecting their bicriterion vectors to begin the recursion. However, for the tricriterion case, this may cause difficulties due to tiebreaking. Continuing with the example in Table 1, if one initializes the algorithm with paths P1, P2, and P4 (least-mean time, least variance, and least-cost paths, respectively), path P3 (which is also a least variance path) will never be identified. This difficulty is avoided by creating three artificial tricriterion vectors $(M, 0, 0)$, $(0, M, 0)$, and $(0, 0, M)$, for a large scalar M , and using their convex hull as the initial plane segment.

Once the efficient paths have been generated, travelers must be assigned to each path based on their values of travel time and reliability. The number of travelers choosing a path is equal to the double integral of the joint density functions for $VOATT$ and $VOTR$, taken over the corresponding region in the dual graph. If the density function is difficult to integrate, an alternate method is to use a Monte Carlo method to generate points $(VOTT, VOTR)$, and applying a point-in-polygon algorithm (see, for instance, Preparata and Shamos, 1985, pp. 41–67) to identify the appropriate path.

Given (fixed) tolls, mean travel times, and travel time variances, this procedure can be used to assign trips to users with varying values of time and reliability. The equilibrium algorithms presented in Dial (1996, 1997) can then apply this assignment procedure repeatedly to find a user equilibrium with heterogeneous users (the remaining modifications to account for three criteria, rather than two, are straightforward). Following Dial (1999a), system-optimal link

flows and first-best tolls can then be found using this equilibrium procedure, by defining link tolls as a function of link flows:

$$\tau_{ij}(x_{ij}) = x_{ij}(\overline{VOATT}_{ij} \frac{dE[t_{ij}]}{dx_{ij}} + \overline{VOTR}_{ij} \frac{dVar[t_{ij}]}{dx_{ij}}) \quad (26)$$

where \overline{VOATT}_{ij} and \overline{VOTR}_{ij} are the mean values of average travel time and reliability for all users on link (i, j) . With this toll function, a tricriterion user equilibrium coincides with the system-optimum.

By accounting for variation in user preferences, the model more accurately represents traveler decision-making, and thus allows more accurate selection of toll levels.

5.3 Elastic Demand

Demand for travel is not independent of travel cost — rather, elasticity is observed as increasing travel costs result in decreased demand. Accounting for this is especially important in pricing applications, where travel costs may be substantially raised relative to the “no-toll” scenario. This can be represented by replacing the fixed OD demands d_{rs} with functions $d_{rs} = \psi_{rs}(\kappa_{rs})$ mapping the cost κ_{rs} of the cheapest r - s path to the demand for travel between r and s . For independent arc states, the system-optimal traffic assignment problem with elastic demand can be written

$$\min_{\mathbf{x}} \sum_{(i,j) \in A} x_{ij} (VOATT \times E[\tilde{t}_{ij}] + VOTR \times Var[\tilde{t}_{ij}]) - \sum_{(r,s) \in D} \int_0^{d_{rs}} \psi_{rs}^{-1}(z) dz \quad (27)$$

$$\text{s.t.} \quad \sum_{\pi \in \Pi_{rs}} h^\pi = d_{rs} \quad \forall (r, s) \in D \quad (28)$$

$$h^\pi \geq 0 \quad \forall \pi \in \Pi \quad (29)$$

$$d_{rs} \geq 0 \quad \forall (r, s) \in D \quad (30)$$

for the no-information scenario, following a well-known formulation of this problem (see, for instance, Sheffi, 1985). The full-information scenarios can be represented by setting $Var[\tilde{t}_{ij}]$ to zero for all links, and $VOATT = VOTT$. From the objective function, it should be clear that a set of link flows and OD demands solving this program maximizes consumer surplus. Once the system-optimal link flows and OD demands are known, the appropriate toll vector can be found using the methods described in the previous sections with the demand fixed at the optimal levels.

6 Application of the Basic Model

The impacts of different information scenarios and users’ valuation of reliability were studied using the well-known Sioux Falls test network, containing 24 nodes,

76 links, and assigning 360,600 vehicle trips (Figure 4), as obtained from Bar-Gera (2007). The basic model developed in Sections 3 and 4 is applied — in the interest of space and clarifying the main impacts of uncertain supply, the extensions in the previous section are not considered here. The capacity on freeway links was made random, equal to its nominal value with probability 0.90, and reduced to a third of its nominal value with probability 0.10 (representing a major incident occurring one day out of ten). A \$10/hr value of travel time was assumed, for both *VOTT* and *VOATT*.

In the absence of data to calibrate the *VOTR* parameter directly, a rough estimate is made based on the results of Small et al. (2005), whose revealed preference data showed travelers were willing to spend \$19.56/hr to reduce the difference between 80th- and 50th-percentile travel times, compared to a \$21.46/hr value of travel time; with our *VOTT* assumption, one expects that our travelers would pay \$9.11 for the same, the proportionate amount. Since their experiment most closely resembled the NI/UT scenario (although tolls changed dynamically, this was done in response to recurrent congestion, rather than incidents), the basic model was initially run with $VOTR = 0$ and the average travel time found. Assuming a normal distribution on trip travel time, the difference between the 80th- and 50th-percentile travel times for the NI/UT scenarios is 1.46 minutes, indicating that our travelers would pay \$0.22 to eliminate this uncertainty; with a variance of 2.97 minutes squared, this implies a \$0.074/min² *VOTR*.

A sample of fifty network realizations was drawn, and each of the solution methods in the previous section was applied to the appropriate information scenario. For comparison with heuristics H1 and H2 for the FI/UT scenario, the simulated annealing (SA) metaheuristic, developed by Kirkpatrick et al. (1983), was also applied to generate an approximately optimal toll vector; for SA, solution neighbors were obtained by perturbing arc tolls by up to fifty cents each. Additionally, for comparison, “no-toll” scenarios were evaluated for the NI and FI user information scenarios. Note that the NI and FI scenarios are still commensurable despite the difference in cost functions, since the FI cost function is simply a special case of the NI cost function with zero variance in travel times.

Tables 2 and 3 compare the mean and standard deviation of trip durations and toll charges under the different information scenarios, along with the computation time needed for each solution method. Table 2 compares “average” trip characteristics; that is, the mean and variance of each traveler’s day-to-day travel times and tolls were first calculated, and averages of these values were taken across all travelers. Table 3 also lists the numerical value of the manager’s objective function, representing the total burden due to travel time and travel variability. Note that the standard deviations shown represent the variation seen over a period of many days — although travelers experience no uncertainty within a given day for the fully informed cases, there still is variation between days in their experienced travel times and toll expenses.

Several observations are apparent. First, providing users with pre-trip information on system conditions provides a substantial reduction in average travel

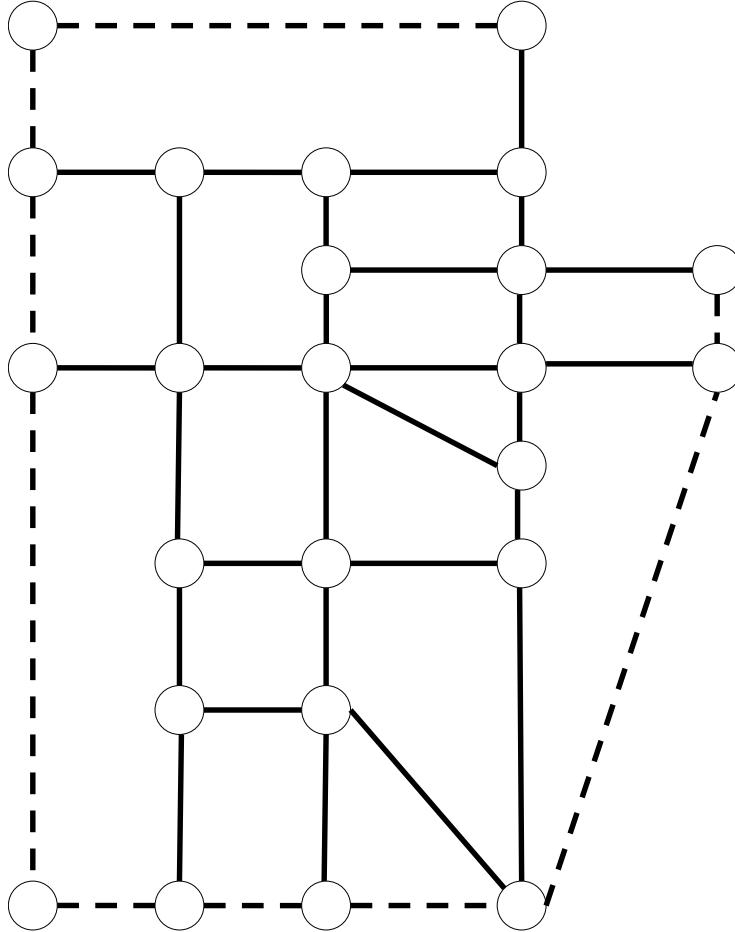


Figure 4: Sioux Falls test network; dashed lines indicate degradable (freeway) links.

Table 2: Comparison of average trip characteristics for all information scenarios

Users	Operator	Travel time (min)		Toll paid (\$)		Run time (s)
		Mean	Std. Dev.	Mean	Std. Dev.	
NI	No Toll	23.60	0.74	0	0	—
	UT	23.55	0.70	9.82	0	5
FI	No Toll	21.51	0.87	0	0	—
	RT	20.15	0.48	6.83	0.01	702
	UT - H1	20.21	0.60	6.86	0.01	385
	UT - H2	20.21	0.60	6.86	0.01	3446
	UT - SA	20.21	0.60	6.86	0.01	891

Table 3: Comparison of system states for all information scenarios

Users	Operator	Total travel time (veh-hr)		Toll revenue ($\$ \times 10^3$)		Objective ($\$ \times 10^3$)
		Mean	Std. Dev.	Mean	Std. Dev.	
NI	No Toll	141851	4418	0	0	1433
	UT	141558	4198	3542	0	1429
FI	No Toll	129286	5253	0	0	1313
	RT	121093	2091	2462	110	1217
	UT - H1	121450.0	3588.7	2474	797	1224
	UT - H2	121449.8	3589.0	2474	799	1224
	UT - SA	121450.3	3588.7	2474	794	1224

times, on the order of ten to fifteen percent. Second, marginal-cost tolls are higher when users do not have access to information on the network state. This occurs because the effect of a potential incident must always be incorporated into the toll price — since an incident result in large delays, a large toll is needed to correct the situation. Responsive tolling and providing users information allow more finesse: if users are aware of an incident, many will choose alternate routes on their own, even without a high toll; and responsive tolling allows levying a high toll only when warranted by an incident. A more mathematical reason is the Pigouvian toll must include a term representing the marginal loss in reliability in addition to the marginal increase in average travel time, unless information is provided.

Third, congestion pricing has a greater impact in improving average travel times (as compared to the no-toll case) when users have information, but the marginal-cost tolls are much higher on average. For the responsive tolling scenarios, this occurs because the tolls on degraded links can be selectively increased, providing additional disincentive for using such links — without tolling, the increased travel times also discourage use of these links, but prices allow for an even greater reduction in total system travel time. Even for the unresponsive tolling scenarios, high tolls appear to be needed, perhaps to prevent users from “overcorrecting” when they learn of reduced capacity on their original path choice, creating additional congestion on a secondary route even as their own travel time decreases.

Finally, as is common with marginal-cost tolling, the levied tolls are greater in magnitude than the reduction in travel disutility. Nevertheless, since toll revenues are assumed to be returned to travelers in some fashion, as long as the cost of implementing and administering the toll system is smaller than the reductions in disutility indicated in Table 3, a net social benefit still obtains.

For finding tolls in the FI/UT case, simple averaging (H1) appears to work just as well as the alternate heuristic H2, or simulated annealing, while requiring far less computation time; it would be interesting to see whether this result occurs in larger networks as well.

More insight is obtained by examining differences among individual link conditions under these different scenarios. As base cases for comparison, Figures 5

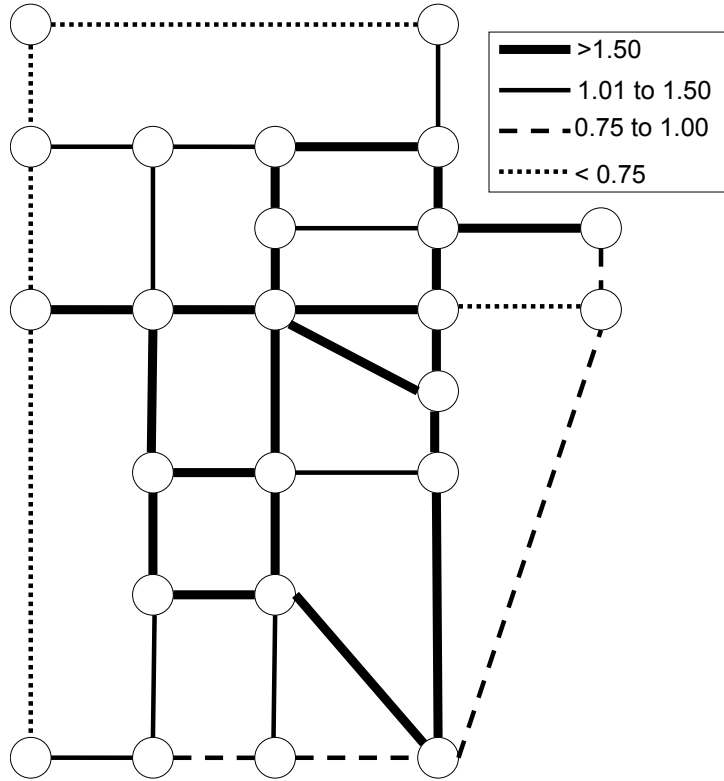


Figure 5: Untolled volume-to-capacity ratios for no-information case

and 6 respectively show average link volume-to-capacity (v/c) ratios for the no-information and full-information cases when no tolls are present. Note that although arterials are substantially congested (almost all have $v/c > 1$), freeway links are generally underutilized, as a result of their uncertain capacity. When travelers have information, freeways are used more, especially in the southern part of the network where congestion becomes apparent. Higher freeway usage is also present in the northern part of the network, but spare capacity remains.

Figures 7, 8, and 9 show the change in v/c ratios when tolls are applied in the NI/UT, FI/RT, and FI/UT cases, compared with the untolled base cases. For the uninformed (NI/UT) scenario, the general effect of the tolls is to shift flow onto less congested routes, such as the northern freeways. Increased freeway usage is especially apparent in the informed (FI/RT and FI/UT) scenarios, where the combination of information provision and tolls are effective in persuading travelers to use less-congested freeways to the north and west, leading to gains in system-wide operations.

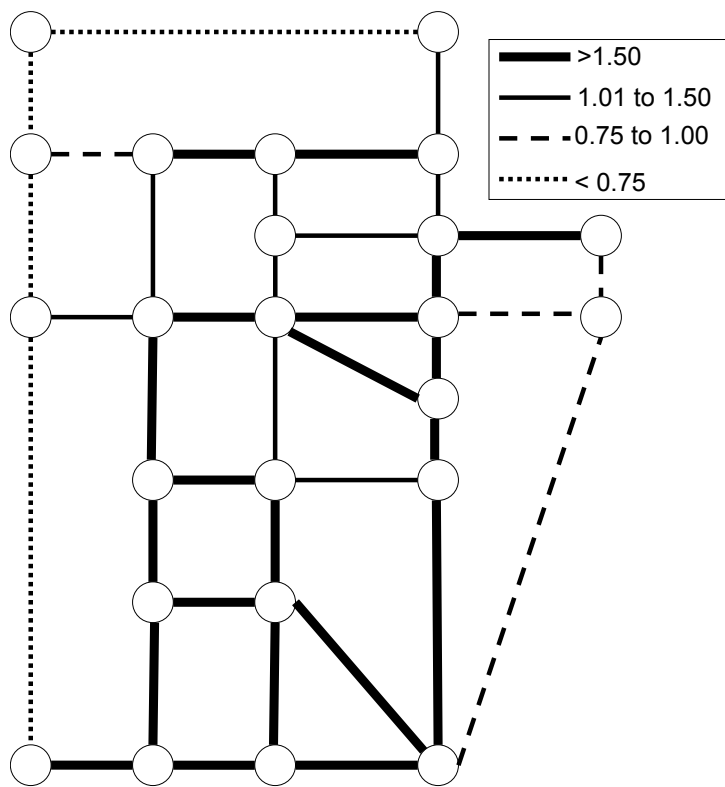


Figure 6: Untolled volume-to-capacity ratios for full-information case

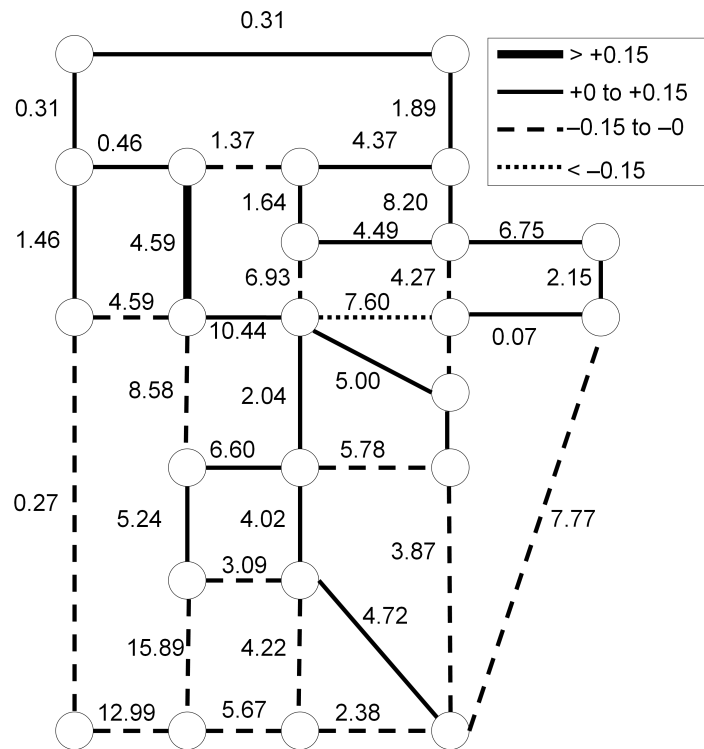


Figure 7: Change in v/c ratios under tolling (NI/UT); average link tolls shown in dollars

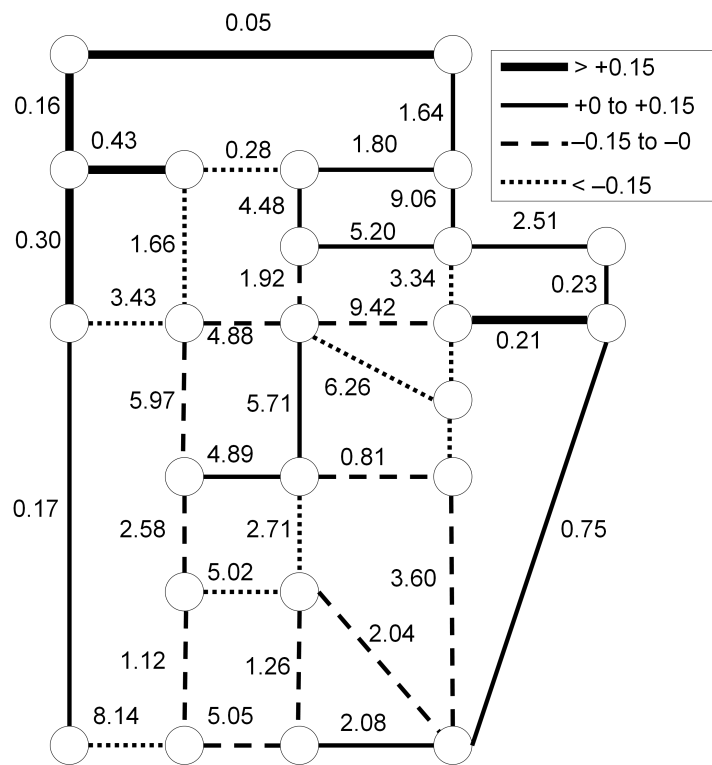


Figure 8: Change in v/c ratios under tolling (FI/RT); average link tolls shown in dollars

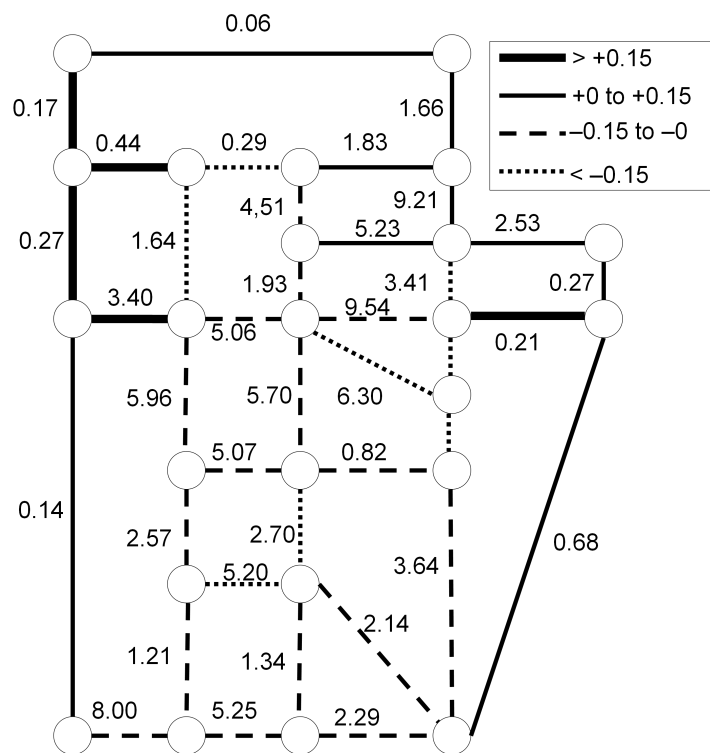


Figure 9: Change in v/c ratios under tolling (FI/UT); average link tolls shown in dollars

7 Conclusion

This paper considers first-best pricing problems in the presence of network uncertainty and user valuation of travel time reliability. As with any stochastic model, the question of information is key. Four information provision scenarios are developed, accounting for both network managers and users, although one of these is shown to be a special case of another: from the standpoint of encouraging system-optimal behavior, there is no value in varying tolls if users do not learn of the network realization. Solution methods are presented for each of these scenarios, which were then tested on the Sioux Falls network.

Perhaps the most noteworthy conclusion is that unresponsive tolls must be set higher than responsive tolls, since the network manager must always hedge against rare events to ensure the convergence of system optimal and user optimal behavior. This suggests that unresponsive tolling should not be used to address nonrecurring congestion, but instead be limited to recurring, predictable congestion. Responsive tolls, however, do not suffer from this weakness, assuming full information on behalf of network managers and travelers.

There are many interesting extensions to this work that are worthy of future study. For instance, the use of a network including real data on capacity variations, with the *VOTR* parameter estimated directly from traveler choices, rather than estimated from another model, would do much to illuminate whether the numerical results can be generalized. From a theoretical standpoint, perhaps the best way to account for uncertainty is to allow adaptive route choice, where information on tolls and network conditions is learned while a trip is in progress, and accounting for this would greatly increase the realism of this model. Further, additional information scenarios can be studied, including the possibility of imperfect or costly information. Finally, perhaps the most significant limitation of the current approach is its static nature: the advantages of dynamic traffic assignment are well-known, and are capable of modeling network disruptions with greater realism. All of these extensions are of interest for further explorations of the problem of network pricing under uncertainty.

References

- Akamatsu, T. and M. Kuwahara (1988). Optimal road pricing under stochastic user equilibrium. *Proceedings of the Japan Society of Civil Engineers 389/IV-8*, 121–129.
- Avineri, E. and J. N. Prashker (2003). Sensitivity to uncertainty: Need for a paradigm shift. *Transportation Research Record 1854*, 90–98.
- Bar-Gera, H. (2007). Transportation test problems. Website: <http://www.bgu.ac.il/~bargera/tntp/>. Accessed November 17, 2007.
- Bates, J., J. Polak, P. Jones, and A. Cook (2001). The valuation of reliability for personal travel. *Transportation Research Part E 37*(2–3), 191–229.

- Bertsimas, D. and J. N. Tsitsiklis (1997). *Introduction to Linear Optimization*. Cambridge, MA: Athena Scientific.
- Brownstone, D. and K. A. Small (2005). Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A* 39(4), 279–293.
- Casella, G. and R. L. Berger (2002). *Statistical Inference* (2nd ed.). Pacific Grove, CA: Duxbury.
- Chan, K. S. and W. H. K. Lam (2005). Impact of road pricing on the network reliability. *Journal of the Eastern Asia Society for Transportation Studies* 6, 2060–2075.
- de Palma, A. and R. Lindsey (1998). Information and usage of congestible facilities under different pricing regimes. *Canadian Journal of Economics* 31(3), 666–692.
- de Palma, A. and N. Picard (2005). Route choice decision under travel time uncertainty. *Transportation Research Part A* 39, 295–324.
- Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research* 5, 83–111.
- Dial, R. B. (1996). Bicriterion traffic assignment: basic theory and elementary algorithms. *Transportation Science* 30, 93–111.
- Dial, R. B. (1997). Bicriterion traffic assignment: efficient algorithms plus examples. *Transportation Research Part B* 31(5), 357–379.
- Dial, R. B. (1999a). Network-optimized road pricing: Part I: a parable and a model. *Operations Research* 47(1), 54–64.
- Dial, R. B. (1999b). Network-optimized road pricing: Part II: algorithms and examples. *Operations Research* 47(2), 327–336.
- Emmerink, R. H. M., E. T. Verhoef, P. Nijkamp, and P. Rietveld (1996). Endogenising demand for information in road transport. *Annals of Regional Science* 30(2), 201–222.
- Eswaran, P. K., A. Ravindran, and H. Moskowitz (1989). Algorithms for non-linear integer bicriterion problems. *Journal of Optimization Theory and Applications* 63(2), 261–279.
- Falk, M. (1999). A simple approximation to the generation of uniformly distributed random variables with prescribed correlations. *Communications in Statistics — Simulation and Computation* 28(3), 785–791.
- Fudenberg, D. and D. K. Levine (1998). *The Theory of Learning in Games*. Cambridge, MA: MIT Press.

- Gabriel, S. A. and D. Bernstein (2000). Nonadditive shortest paths: subproblems in multi-agent competitive network models. *Computational and Mathematical Organization Theory* 6(1), 29–45.
- Gao, S. (2005). *Optimal adaptive routing and traffic assignment in stochastic time-dependent networks*. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Geoffrion, A. M. (1968). Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications* 22, 618–630.
- Haimes, Y. Y., L. S. Lasdon, and D. A. Wismer (1971). On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, and Cybernetics* 1(3), 296–297.
- Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983). Optimization by simulated annealing. *Science* 220(4598), 671–680.
- Kobayashi, K. and M. Do (2005). The informational impacts of congestion tolls upon route traffic demands. *Transportation Research Part A* 39, 651–670.
- Labbé, M., P. Marcotte, and G. Savard (1998). A bilevel model of taxation and its application to optimal highway pricing. *Management Science* 44(12), 1608–1622.
- Lindsey, R. (2008). Cost recovery from congestion tolls with stochastic capacity and demand. Presented at the Third International Conference on Funding Transport Infrastructure, Paris, France.
- Liu, H. X., X. He, and W. Recker (2007). Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research Part B* 41(4), 448–461.
- Liu, H. X., W. Recker, and A. Chen (2004). Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. *Transportation Research Part A* 38(6), 435–453.
- Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance* 7(1), 77–91.
- Mohring, H. and M. Harwitz (1962). *Highway Benefits: An Analytical Framework*. Evanston, IL: Northwestern University Press.
- Noland, R. B. and J. W. Polak (2002). Travel time variability: a review of theoretical and empirical issues. *Transport Reviews* 22(1), 39–54.
- Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan and Co.
- Pinjari, A. and C. Bhat (2006). On the nonlinearity of response to level of service variables in travel mode choice models. *Transportation Research Record* 1977, 67–74.

- Polychronopoulos, G. H. and J. N. Tsitsiklis (1996). Stochastic shortest path problems with recourse. *Networks* 27(2), 133–143.
- Preparata, F. P. and M. I. Shamos (1985). *Computational Geometry*. New York, NY: Springer-Verlag.
- Sheffi, Y. (1985). *Urban Transportation Networks*. Englewood Cliffs, NJ: Prentice-Hall.
- Small, K. A., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica* 73(4), 1367–1382.
- Smith, T. E., E. A. Eriksson, and P. O. Lindberg (1994). Existence of optimal tolls under conditions of stochastic user-equilibria. In B. Johansson and L. G. Mattsson (Eds.), *Road Pricing: Theory, Empirical Assessment and Policy*, pp. 65–87. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Sternbach, M. C. (2001). Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Review* 43(1), 31–85.
- Supernak, J., C. Kaschade, and D. Steffey (2003). Dynamic value pricing on I-15 in San Diego: Impact on travel time and its reliability. *Transportation Research Record* 1839, 45–54.
- Verhoef, E. T., R. H. M. Emmerink, P. Nijkamp, and P. Rietveld (1996). Information provision, flat and fine congestion tolling and the efficiency of road usage. *Regional Science and Urban Economics* 26, 505–529.
- Waller, S. T. and A. K. Ziliaskopoulos (2002). On the online shortest path problem with limited arc cost dependencies. *Networks* 40(4), 216–227.
- Yang, H. (1999a). Evaluating the benefits of a combined route guidance and road pricing system in a traffic network with recurrent congestion. *Transportation* 26, 299–322.
- Yang, H. (1999b). System optimum, stochastic user equilibrium, and optimal link tolls. *Transportation Science* 33(4), 354–360.