The HKU Scholars Hub The University of Hong Kong 香港大學學術庫

Title	Environmentally sustainable toll design for congested road networks with uncertain demand
Author(s)	Li, ZC; Lam, WHK; Wong, SC; Sumalee, A
Citation	International Journal Of Sustainable Transportation, 2012, v. 6 n. 3, p. 127-155
Issued Date	2012
URL	http://hdl.handle.net/10722/150605
Rights	This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Sustainable Transportation on 03 Oct 2011, available online: http://www.tandfonline.com/doi/abs/10.1080/15568318.2011.570101

Environmentally sustainable toll design for congested road networks with uncertain demand

Zhi-Chun Li,[a,b] William H.K. Lam,[a,*] S.C. Wong,[c] A. Sumalee[a]

[a]*Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China*

[b]*School of Management, Huazhong University of Science and Technology, Wuhan 430074, China*

[c]*Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China*

Abstract

This paper proposes a new road toll design model for congested road networks with uncertain demand that can be used to create a sustainable urban transportation system. For policy assessment and strategic planning purposes, the proposed model extends traditional congestion pricing models to simultaneously consider congestion and environmental externalities due to vehicular use. Based on analyses of physical and environmental capacity constraints, the boundary conditions under which a road user on a link should pay either a congestion toll or an extra environmental tax are identified. The sustainable toll design model is formulated as a two-stage robust optimization problem. The first-stage problem before the realization of the future travel demand aims to minimize a risk-averse objective by determining the optimal toll. The second stage after the uncertain travel demand has been determined is a scenario-based route choice equilibrium formulation with physical and environmental capacity constraints. A heuristic algorithm that combines the sample average approximation approach and a sensitivity analysis-based method is developed to solve the proposed model. The upper and lower bounds of the model solution are also estimated. Two numerical examples are given to show the properties of the proposed model and solution algorithm and to investigate the effects of demand variation and the importance of including risk and environmental taxation in toll design formulations.

Keywords: Sustainable toll design; congestion and environmental externalities; uncertain demand; two-stage robust optimization; capacity constraint; sample average approximation.

[*]Corresponding author. Tel.: +852-2766-6045; fax: +852-2334-6389.
E-mail address: cehklam@polyu.edu.hk (W.H.K. Lam).

1. Introduction

Global warming, or climate change, is one of the most serious threats facing the world today, and considerable attention has been paid to sustainable environmental issues. There is a broad consensus that transportation systems are major contributors to climate change due to various externalities, including congestion and environmental impacts (Taylor 1996; Nagurney 2000a; Black and Sato 2007). Studies show that vehicular use contributes to 30-50% of hydrocarbon, 40-60% of nitrogen oxide, and 80-90% of carbon monoxide emissions (United States Environmental Protection Agency 1991, 1992). The Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC 2007) stated that, during the 1990s, carbon emissions increased by less than 1% per year, whereas since 2000, emissions have grown at a rate of 3.5% per year, exacerbating environmental problems. There is clearly an urgent need for effective measures and policies to combat further environmental damage due to increased vehicular pollution emissions and to develop sustainable, low-carbon urban transportation systems.

Road toll pricing, which is a type of traffic demand management measure, is widely recognized as a useful tool for alleviating traffic congestion and reducing vehicular emissions. Recent developments in information and communication technologies have made the implementation of road pricing schemes easier. Well-known successful examples of electronic road pricing include congestion-charging schemes in California, Singapore, and, more recently, London (Santos 2004; Hau 2006).

There is a substantial body of literature on road toll pricing issues. The approaches adopted in previous related studies can generally be classified into two main categories: first-best pricing and second-best pricing. First-best pricing is based on the fundamental economic principle of marginal cost pricing, which requires that road users on each link of a network pay a toll equal to the difference between the marginal social and marginal private costs. However, this ideal pricing scheme may not be applicable in reality due to political and social restrictions. This has motivated research into the second-best pricing, where only a subset of links can be subject to tolls. A thorough discussion of the economic fundamentals of road pricing has been given by Hau (2005) and Yang and Huang (2005), and a more recent comprehensive review can be found in Tsekeris and Voß (2009).

Various studies have been carried out to develop different models (deterministic or stochastic) to assess the effects of different types of road toll pricing (congestion or environmental). Table 1 gives a summary of these studies and reveals that most mainly dealt with the extra time costs (i.e., congestion externalities) that each road user imposes on others, which are collectively referred to as congestion pricing. In contrast, environmental externalities caused by vehicular traffic have received little attention. Rouwendal and Verhoef (2006) argued that ignoring environmental externalities in road toll pricing can lead to market failure. Other studies, such as those of Nagurney (2000b), Yin and Lawphongpanich (2006), and Szeto, Li, and O'Mahony (2008), have shown that the implementation of congestion pricing can lead to an emissions paradox whereby traffic emissions actually increase. It is thus necessary to incorporate environmental externalities into road toll pricing models to achieve environmentally sustainable urban transportation systems, particularly in this era of climate change.

Table 1 also indicates that existing road toll pricing studies mainly use deterministic models, and that little attention has been paid to stochastic solutions. However, in reality, traffic flows on urban roads are basically stochastic. This stochastic effect may be due to various random factors that range from irregular and random incidents, such as traffic accidents, vehicle breakdowns, road work, signal failures, adverse weather, and earthquakes, to regular fluctuations in travel demand and capacity by time of day, day of week, and season (Chen et al. 2002, 2010; Chen, Subprasom, and Ji 2003; Chootinan, Wong, and Chen 2005; Lo, Luo, and Siu 2006; Li et al. 2008; Yin, Madanat, and Lu 2009; Kim, Kurauchi, Uno 2011). As a result, the optimal tolls derived using deterministic models may negatively affect the performance of transportation systems, especially when the actual future demand and supply deviate significantly from their expected values (Gardner, Unnikrishnan, and Waller 2008). It is thus important to incorporate the effects of uncertain demand and supply into road toll pricing models for policy assessment.

In view of the shortcomings of previous studies in this area, this study proposes an environmentally sustainable toll design model for congested road networks with demand uncertainty. It extends previous work as follows. (1) A new robust toll design model that explicitly incorporates the effects of demand uncertainty and congestion and environmental externalities is developed for strategic planning. The proposed modeling framework has

important implications for the development of sustainable urban transportation systems. (2) Based on analyses of the relationship between physical and environmental capacity constraints, the boundary conditions under which a road user on a link should pay either a congestion toll or an extra environmental tax for vehicular use are determined. (3) A heuristic algorithm that is a combination of the sample average approximation approach and a sensitivity analysis-based method is developed to solve the proposed model. The quality (lower and upper bounds) of the model solution is also examined together with the efficiency of the proposed solution algorithm and the effects of variation in pricing schemes and demand.

The proposed model consists of two decision stages and two groups of decision variables, one in each stage. As shown in Figure 1, the first stage of the model, which is referred to as the "here-and-now" stage, occurs before the actual travel demand is realized. The performance of the transport system, including the total user travel time and total amount of traffic emissions, is a random variable and is highly dependent on the actual future travel demand. The optimal toll solution obtained by traditional expectation models, which optimize the expected value of an objective function over all possible demand scenarios, can lead to large variation in the objective values for different scenarios. In particular, some scenarios may have very poor objective values. To compensate for this limitation of expectation models, this study introduces a risk-averse objective into the first stage of the proposed robustness model. This objective is defined as a linear bi-criteria combination of the expected value of an objective function and its semi-deviation or mean absolute deviation. The second stage of the model, which is referred to as the "wait-and-see" stage, models the responses (in terms of route choice) of road users to the first-stage toll decisions after a particular travel demand scenario has been realized. For a given demand scenario, each road user in the network is assumed to select the route with the minimal travel disutility, which leads to the classical Wardrop user equilibrium.

The remainder of this paper is organized as follows. The next section describes the basic assumptions and capacity constraints. Section 3 presents the two-stage robust optimization formulation. Section 4 presents the development of a heuristic solution algorithm for solving the proposed model and discusses the quality of the solution. Section 5 provides two examples to illustrate the application of the proposed model and solution algorithm. The conclusion is given in Section 6 together with recommendations for further studies.

2. Basic considerations

2.1. Assumptions

To facilitate the presentation of the essential ideas of this paper, the following assumptions are made.

A1 The implementation of road tolls aims to simultaneously restrain the levels of traffic congestion and vehicular emissions within acceptable limits. To do so, two categories of capacity constraints – physical and environmental (emissions) – are introduced (Ferrari 1997; Yang and Bell 1997; Chen, Zhou, and Ryu 2011).

A2 The travel demand between each origin-destination (OD) pair is a random variable with a known probability distribution that can be discretized as a finite set of demand scenarios or realizations (Gardner, Unnikrishnan, and Waller 2008; Yin, Madanat, and Lu 2009; Boyles, Kockelman, and Waller 2010). The objective of the road toll design model developed in this study is to minimize the linear bi-criteria combination of the expected value of the total system travel disutility (i.e., total user travel time plus total emissions) and its semi-deviation risk measure over all demand scenarios. For each demand scenario, the route choice of road users follows the Wardrop user equilibrium principle, but is subject to physical and environmental capacity constraints.

A3 Carbon monoxide (CO) is considered to be an indicator of the level of atmospheric pollution generated by vehicular traffic. This assumption is not unreasonable because (i) vehicles are responsible for almost all CO emissions in the air, (ii) CO is the most significant pollutant among the various types of vehicular emissions, which include nitrogen oxide, CO, nitrogen dioxide, sulfur dioxide, ozone, and particulates, and (iii) the emission rates or production functions of other pollutants are similar to that of CO. This assumption has also been made in related studies, including those of Rilett and Benedek (1994), Benedek and Rilett (1998), Wallace et al. (1998), Sugawara and Niemeier (2002), Yin and Lawphongpanich (2006), and Nagurney, Qiang, and Nagurney (2010).

A4 The study period is assumed to be a one-hour period, such as the morning peak hour, which is usually the most critical period in the day. The model proposed here is thus mainly for the purposes of strategic planning or policy evaluation. The vehicular pollutant emission

rate in the study period is estimated by the macroscopic model suggested by Penic and Upchurch (1992), and is related to the average vehicle speed (including the constant, acceleration, and deceleration rates) and delays at intersections due to signal controls, if any.

A5 This study mainly focuses on automobiles, and other types of vehicles (e.g., buses) are not considered. The (average) travel time of an automobile on each link of the road network is assumed to be a continuous and strictly increasing function of the flows on that link, which implies that the link interaction is not considered. All of the users in the network are assumed to be homogeneous in terms of their time value. However, this assumption can easily be relaxed to consider multiple transport modes, multiple user classes, or the interaction of flows across links.

2.2. Physical capacity constraint

Physical capacity constraints are inherent in road transport systems, and there is a high risk of system failure or flow instability if they are violated (Ferrari 1995, 1997). Consider a road network $G = (N, A)$ where N is the set of all nodes and A is the set of all links in the network. Let W be the set of all origin-destination (OD) pairs in the network and R_w be the set of all routes between OD pair $w \in W$. Let Ω be the set of all demand scenarios (or realizations) and s a particular demand scenario $s \in \Omega$. Let v_a^s be the flow on link a under demand scenario s. The physical capacity constraint can then be represented as

$$v_a^s \le C_a, \quad \forall a \in A, s \in \Omega, \tag{1}$$

where C_a is the physical capacity of link a, $a \in A$, as measured in vehicles per hour. This may be interpreted as the exit capacity of a link because the capacity at an intersection is usually minimal. For a signal-controlled intersection, the exit capacity of a link approaching the intersection depends on the proportion of green time, and can be calculated using the following formula (see Yang and Yagar 1995; Wong and Yang 1997).

$$C_a = \rho_a \bar{C}_a, \quad \forall a \in A, \tag{2}$$

where ρ_a $(0.0 < \rho_a \le 1.0)$ is the proportion of a cycle that is effectively green for link a, and $\bar{C}_a$ is the (constant) saturation flow (in vehicles per hour) of link a. When a link approaches a signal-controlled intersection, the value of ρ_a is between 0 and 1, that is, $0 < \rho_a < 1$. When there is no signal control at the exit of a link, the value of ρ_a is equal to 1,

that is, $C_a = \bar{C}_a$.

The link flow v_a^s in Equation (1) can be given by

$$v_a^s = \sum_{w \in W} \sum_{r \in R_w} f_{rw}^s \delta_{ar}, \quad \forall a \in A, s \in \Omega, \tag{3}$$

where f_{rw}^s is the flow on route $r \in R_w$ between OD pair $w \in W$ under demand scenario $s \in \Omega$. δ_{ar} equals 1 if route *r* traverses link *a*, and 0 otherwise.

2.3. Environmental capacity constraints

Environmental (or emission) capacity constraints are not inherent in road networks, but are imposed by the system manager (e.g., the authority) on some links of the network to keep the level of local traffic emissions below a certain threshold (such as the maximum permissible emission standard). Let $e_a^s(v_a^s)$ be the average amount of traffic emissions on link *a* under demand scenario *s*, which is measured in grams per vehicle. According to **A4**, $e_a^s(v_a^s)$ consists of two components: emissions $e_a^{\mathrm{s,Move}}$ due to vehicular movements on link *a*, and emissions $e_a^{\mathrm{s,Delay}}$ due to a stop delay at a signal-controlled intersection (if any), that is,

$$\begin{aligned} e_a^s(v_a^s) &= e_a^{\mathrm{s,Move}} + e_a^{\mathrm{s,Delay}}, \\ &= l_a \times \left[\frac{9.1913}{\bar{k}_a^s} \exp\left(0.01023 \bar{k}_a^s\right) \right] + 0.003\left(1-\rho_a\right)\Gamma, \quad \forall a \in A, s \in \Omega, \end{aligned} \tag{4}$$

where the term in the square brackets on the right-hand side of the equation represents the average amount of traffic emissions per vehicle-kilometer. l_a is the length (in kilometers) of link *a*, and $\bar{k}_a^s$ is the average vehicle speed (in kilometers per hour) on link *a* under scenario *s*. When converted to feet and feet per second, the numbers in the square brackets are the same as those given by Penic and Upchurch (1992). Γ is the constant signal cycle time, measured in seconds, and $\left(1-\rho_a\right)\Gamma$ is the average stop delay (also in seconds) of vehicles on link *a* due to signal control. The calibrated coefficient 0.003 is the delay emissions per vehicle-second (Penic and Upchurch 1992). When a link is not subject to signal control, ρ_a equals 1, which means that the second term on the right-hand side of Equation (4) is zero.

Let $t_a^s(\cdot)$ be the average travel time (in hours) of vehicles on link *a* under scenario *s*, that is, $t_a^s = l_a / \bar{k}_a^s$ (not including a stop delay at a signal intersection). Equation (4) can then be expressed as

$$e_a^s(v_a^s) = 9.1913 \cdot t_a^s(v_a^s) \cdot \exp\left(0.01023 \frac{l_a}{t_a^s(v_a^s)}\right) + 0.003(1-\rho_a)\Gamma, \quad \forall a \in A, s \in \Omega, \qquad (5)$$

where the link travel time function $t_a^s(v_a^s)$ is assumed to be continuous and strictly increasing with regards to v_a^s. It can be estimated by the following Bureau of Public Roads (BPR) type function

$$t_a^s(v_a^s) = \frac{l_a}{k_a^0}\left(1 + \alpha\left(\frac{v_a^s}{C_a}\right)^n\right), \quad \forall a \in A, s \in \Omega, \qquad (6)$$

where k_a^0 is the (constant) free-flow speed of vehicles on link *a*, measured in kilometers per hour. The capacity C_a of link *a* can be calculated using Equation (2). The parameters α and *n* reflect congestion effects.

Let $\hat{e}_a^s(v_a^s)$ be the total amount of emissions on link *a* under demand scenario *s*, measured in grams per hour. It is the product of the average amount of traffic emissions and the link flow (the number of vehicles) on that link, and is represented by

$$\hat{e}_a^s(v_a^s) = v_a^s \cdot e_a^s(v_a^s), \quad \forall a \in A, s \in \Omega. \qquad (7)$$

The environmental or emission capacity constraint can thus be expressed as

$$\hat{e}_a^s(v_a^s) \le E_a, \quad \forall a \in A, s \in \Omega, \qquad (8)$$

where E_a is the maximum permissible emission standard of link *a*, measured in grams per hour.

Studies show that when the vehicle speed is lower than 70 kilometers per hour, $e_a^s(\cdot)$ in Equation (5) or (7) is a monotonically increasing (or decreasing) function of flow (or speed) (see Jakkula and Asakura 2009; Yin and Lawphongpanich 2006) and is thus invertible. This means that $\hat{e}_a^s(\cdot)$ is also invertible. We denote the inverse of $\hat{e}_a^s(\cdot)$ as $(\hat{e}_a^s)^{-1}(\cdot)$, and take the inverse on both sides of inequality (8) to obtain

$$v_a^s \le \bar{E}_a, \quad \forall a \in A, s \in \Omega, \tag{9}$$

where $\bar{E}_a$ is the environmental capacity, measured in vehicles per hour, and

$$\bar{E}_a = (\hat{e}_a^s)^{-1}(E_a), \quad \forall a \in A, s \in \Omega. \tag{10}$$

By using the bisection method (Epperson 2007), Equation (10) can be solved and the unique value of the environmental capacity $\bar{E}_a$ can be obtained.

2.4. Relationship between the two kinds of capacity constraints

As noted, the (non-linear) environmental capacity constraint (Equation (8)) can be equivalently expressed as a linear constraint (i.e., Equation (9)), which is similar to the linear physical capacity constraint (1). However, there is a distinct difference between $\bar{E}_a$ and C_a. Environmental capacity $\bar{E}_a$ in Equation (9) is a variable related to the length of link *a* and the travel time (or speed) of vehicles on link *a* (see Equations (4)-(6)), whereas physical capacity C_a in Equation (1) is a constant for a given signal setting. $\bar{E}_a$ and C_a have the following relationship.

Proposition 1. *Given the values of* E_a, C_a, k_a^0, ρ_a, Γ, *and* α, *the physical and environmental capacities satisfy:*

(i) $\bar{E}_a < C_a$, *if and only if* $l_a > \bar{L}_a$,

(ii) $\bar{E}_a = C_a$, *if and only if* $l_a = \bar{L}_a$, *and*

(iii) $\bar{E}_a > C_a$, *if and only if* $l_a < \bar{L}_a$,

where $\bar{L}_a$ *is referred to as the critical length of link a, which can be determined by*

$$\bar{L}_a = \frac{\left(E_a - 0.003 \cdot C_a \cdot (1-\rho_a)\Gamma\right) k_a^0}{9.1913 \cdot C_a \cdot (1+\alpha)} \cdot \exp\left(-0.01023 \cdot \frac{k_a^0}{1+\alpha}\right).$$

Proof. We first prove entry (i). As $\hat{e}_a^s(\cdot)$ is invertible, we have

$$\bar{E}_a < C_a \Leftrightarrow (\hat{e}_a^s)^{-1}(E_a) < C_a \Leftrightarrow E_a < \hat{e}_a^s(C_a), \quad \forall a \in A. \tag{11}$$

From Equations (5)-(7), we obtain

$$\hat{e}_a^s(C_a) = C_a \cdot e_a^s(C_a)$$

$$= 9.1913 \cdot C_a \cdot \frac{l_a}{k_a^0} \cdot (1+\alpha) \cdot \exp\left(0.01023 \cdot \frac{k_a^0}{1+\alpha} \right) + 0.003 \cdot C_a \cdot (1-\rho_a)\Gamma, \quad \forall a \in A. \quad (12)$$

Combining Equations (11) and (12) yields

$$\bar{E}_a < C_a \Leftrightarrow l_a > \frac{\left(E_a - 0.003 \cdot C_a \cdot (1-\rho_a)\Gamma\right) k_a^0}{9.1913 \cdot C_a \cdot (1+\alpha)} \cdot \exp\left(-0.01023 \cdot \frac{k_a^0}{1+\alpha} \right), \quad \forall a \in A. \quad (13)$$

This completes the proof of entry (i). The proofs of entries (ii) and (iii) are similar.

Proposition 1 shows that the environmental capacity $\bar{E}_a$ of link *a* may be greater than, smaller than, or equal to the physical capacity C_a, depending on the length of the link. For illustrative purposes, Figure 2 shows the change in environmental capacity with the length of the link. In this example, the following input parameters are assumed: $C_a = 3000$ (veh/h), $E_a = 1200$ (g/h), $k_a^0 = 48$ (km/h), $\alpha = 0.15$, $n = 4.0$, and $\rho = 1.0$ (no signal control). Figure 2 shows that as the length l_a of link *a* increases, the environmental capacity $\bar{E}_a$ of link *a* monotonically decreases. The critical length $\bar{L}_a$ of link *a* is 1.2 km. Accordingly, if $l_a < 1.2$ km, then $\bar{E}_a > C_a$; if $l_a > 1.2$ km, then $\bar{E}_a < C_a$; and $\bar{E}_a = C_a$ otherwise.

Corollary 1. *Given the values of* E_a, C_a, k_a^0, ρ_a, Γ, *and* α, *the physical and environmental capacity constraints (1) and (9) satisfy the following:*

(i) If $l_a > \bar{L}_a$, *then* $v_a^s \le \bar{E}_a < C_a$,

(ii) If $l_a = \bar{L}_a$, *then* $v_a^s \le \bar{E}_a = C_a$, *and*

(iii) If $l_a < \bar{L}_a$, *then* $v_a^s \le C_a < \bar{E}_a$.

3. Model formulation

As previously stated, the road toll pricing model developed in this study can be formulated as a two-stage decision process. In the first stage, the road tolls before future demand has been realized are optimized, whereas in the second stage, the route choice behavior of road users is modeled once the demand uncertainty is revealed. In the following, we formulate the two-stage decisions.

3.1. Scenario-based capacitated UE problem

According to **A2**, for a given demand scenario, the route choice behavior of road users can be modeled as a Wardrop user equilibrium (UE) formulation, subject to physical and environmental capacity constraints. Assuming separate link travel time functions (see **A5** and Equation (6)), it can be shown that the UE link-flow pattern can be obtained by solving the following mathematical programming problem.

$$\min_{\mathbf{v}} \sum_{a\in A\setminus\bar{A}} \int_0^{v_a^s} \hat{t}_a^s(\omega)d\omega + \sum_{a\in\bar{A}} \int_0^{v_a^s} \left(\hat{t}_a^s(\omega) + \frac{x_a}{\beta} \right) d\omega, \tag{14}$$

subject to

$$\sum_{r\in R_w} f_{rw}^s = q_w^s, \quad \forall w \in W, s \in \Omega, \tag{15}$$

$$v_a^s = \sum_{w\in W} \sum_{r\in R_w} f_{rw}^s \delta_{ar}, \quad \forall a \in A, s \in \Omega, \tag{16}$$

$$v_a^s \le C_a, \quad \forall a \in A, s \in \Omega, \tag{17}$$

$$v_a^s \le \bar{E}_a, \quad \forall a \in A, s \in \Omega, \text{ and} \tag{18}$$

$$f_{rw}^s \ge 0, \quad \forall r \in R_w, w \in W, s \in \Omega, \tag{19}$$

where $\hat{t}_a^s(v_a^s)$ is the travel time on link *a*, which consists of the flow-dependent running time and a stop delay at the signal intersection (if any), or

$$\hat{t}_a^s(v_a^s) = t_a^s(v_a^s) + (1-\rho_a)\Gamma, \quad \forall a \in A, s \in \Omega, \tag{20}$$

where $t_a^s(v_a^s)$ is given by Equation (6). $\bar{A}$ is the subset of toll links and x_a denotes the tolls on link $a \in \bar{A}$. $\mathbf{v}$ is the vector of the link flow and q_w^s is the demand between OD pair *w* under scenario *s*. β is the road user's time value, and thus $1/\beta$ can be used to convert tolls into equivalent time units. Equations (15) and (16) represent the OD flow and link flow conservation constraints, respectively; Equations (17) and (18) are the physical and environmental capacity constraints, respectively; and Equation (19) is the usual non-negativity constraint on the route flow.

Note that the constraint set of the mathematical programming model (14)-(19) is linear, and its objective function is strictly convex with respect to the link flow variables due to the strict monotony of the link travel time function. Thus, the link flow solution to (14)-(19) is unique.

To show the equivalence between the minimization model (14)-(19) and the UE conditions, we derive the first-order optimality conditions of this model as

$$\begin{cases} \sum_{a\in A} \hat{t}_a^s(v_a^s)\delta_{ar} + \frac{1}{\beta}\sum_{a\in \bar{A}} x_a\delta_{ar} + \sum_{a\in A} d_a^s\delta_{ar} + \sum_{a\in A} \tau_a^s\delta_{ar} = c_w^s, \ \text{ if } f_{rw}^s > 0, \\ \sum_{a\in A} \hat{t}_a^s(v_a^s)\delta_{ar} + \frac{1}{\beta}\sum_{a\in \bar{A}} x_a\delta_{ar} + \sum_{a\in A} d_a^s\delta_{ar} + \sum_{a\in A} \tau_a^s\delta_{ar} \geq c_w^s, \ \text{ if } f_{rw}^s = 0, \end{cases} \tag{21}$$

$$\begin{cases} d_a^s = 0, \ \text{ if } v_a^s < C_a, \\ d_a^s \geq 0, \ \text{ if } v_a^s = C_a, \end{cases} \tag{22}$$

$$\begin{cases} \tau_a^s = 0, \ \text{ if } v_a^s < \bar{E}_a, \\ \tau_a^s \geq 0, \ \text{ if } v_a^s = \bar{E}_a, \end{cases} \tag{23}$$

where c_w^s is the minimum generalized travel cost or disutility of travel between OD pair *w* under demand scenario *s*. d_a^s and τ_a^s are the Lagrange multipliers associated with the physical and environmental capacity constraints (17) and (18), respectively.

The travel disutility of a link is a combination of the flow-dependent travel time, any extra queuing delay, any extra environmental tax, and any congestion toll (for toll links only). Equation (21) indicates that, given demand scenario *s*, for each OD pair, the used routes have the minimal travel disutility, and the travel disutility of any unused route is greater than or equal to the minimum. Clearly, the optimal solution to the mathematical programming model (14)-(19) indeed reproduces the user equilibrium of route choice.

Physically, d_a^s and τ_a^s represent the extra penalties (except the toll x_a) incurred by road users who continue to use the physically or environmentally saturated road links to constrain the link flows at or below the binding capacities. However, as stated, physical capacity constraints are intrinsic to road networks, whereas environmental constraints are imposed by an outside authority. This means that when the flow of a link reaches its physical capacity, a queuing delay occurs on that link. This queuing delay can affect the route choice behavior of users, even in the absence of an intervention (or control). However, users will not change their route choice because of emission constraints in the absence of external intervention. Consequently, there is a substantial difference between the implications of d_a^s and τ_a^s. Specifically, d_a^s can be regarded as the equilibrium queuing delay (the extra time due to

insufficient physical capacity) on physically saturated link *a* in demand scenario *s*, whereas τ_a^s can be interpreted as the extra environmental tax imposed on road users who use environmentally saturated link *a* in demand scenario *s*. The extra environmental tax (which is an economic intervention) is introduced to control the link flows within the pollution emission bounds by affecting the route choice behavior of users.

However, the values of d_a^s and τ_a^s (queuing delays and environmental taxes) may not be unique even though the link-flow pattern is unique. For delays or taxes to be unique, it is necessary and sufficient that all of the binding (or active) physical and environmental capacity constraints are linearly independent (Bell 1995; Yang and Bell 1997). In addition, in solving the capacitated problem (14)-(19), the linear capacity constraints (17) and (18) can be incorporated into the objective function (14) using an augmented Lagrangian penalty function approach. Readers are referred to Larsson and Patriksson (1995) and Li et al. (2007) for detailed descriptions of this approach.

Due to this uniqueness issue, Equations (22) and (23) entail that when capacity constraint (17) or (18) becomes binding (or active), extra queuing delays or environmental taxes will be incurred. The sign of the queuing delay or extra environmental tax of a link can be identified by the following proposition.

Proposition 2. *Given the values of* E_a, C_a, k_a^0, ρ_a, Γ, *and* α, *the signs of* d_a^s *and* τ_a^s *can be determined as follows.*

(i) If $l_a > \bar{L}_a$, *then* $d_a^s = 0$ *and* $\tau_a^s \geq 0$,

(ii) If $l_a = \bar{L}_a$, *then* d_a^s *or* $\tau_a^s \geq 0$, *and*

(iii) If $l_a < \bar{L}_a$, *then* $d_a^s \geq 0$ *and* $\tau_a^s = 0$.

Proof. This can be derived directly from Corollary 1 and the complementary slackness conditions (22) and (23).

Proposition 2 shows that when the length (l_a) of a link is greater than the critical length ($\bar{L}_a$), no queuing occurs on that link but an extra environmental tax may be required, and vice versa.

When the length of a link equals the critical length, the physical and environmental capacity constraints become identical, and thus either of them can be removed from the constraint set. As a result, a queuing delay (or extra environmental tax) is required.

Note that the value of τ_a^s is dependent on the demand scenario, which implies that the environmental tax may vary with different demand scenarios. In order to ensure that the environmental capacity constraint is satisfied for each demand scenario, the environmental tax can take the maximum value of τ_a^s over all demand scenarios, i.e. $\tau_a^* = \max\{\tau_a^s, \forall s \in \Omega\}$.

Thus far, the scenario-based capacitated UE problem has been well defined. The basic underlying idea is to model the effects of demand uncertainty through a finite number of discrete demand scenarios. As such, there is a significant distinction between the proposed scenario-based capacitated UE problem and previous related studies involving demand uncertainty, such as the reliable stochastic user equilibrium (RSUE) problem recently suggested by Sumalee, Watling, and Nakayama (2006), and Lam, Shao, and Sumalee (2008). Specifically, in the scenario-based formulation, all of the constraint conditions must be satisfied for *each* demand scenario. That is, no constraint violation is allowed. However, in the RSUE problem, the flow conversation constraints are defined in terms of an expectation level, and the capacity constraints can be violated at a certain probability level.

3.2. Robust toll design problem

As stated in the introduction section, traditional expectation models may lead to large variations in the performance of a transport system under different demand scenarios. To create a reliable and robust transport system, we introduce a risk-averse measure as the objective of the road toll design.

Let Z^s be the total system disutility under demand scenario *s*, which is expressed as

$$Z^s\left(\mathbf{x}, \mathbf{v}(\mathbf{x}), \mathbf{d}(\mathbf{x})\right) = \beta \sum_{a \in A} v_a^s \left(\hat{t}_a^s(v_a^s) + d_a^s\right) + \gamma \sum_{a \in A} v_a^s e_a^s(v_a^s), \tag{24}$$

where the symbols in bold represent the vectors of the corresponding variables, β is the user's time value (see also Equation (14)), and γ is the damage value per unit of CO emissions. The calibration results in the study of Roth and Ambs (2004) show that the value

of γ is between 506 and 2494 (US$/ton). Further, the congestion toll x_a and extra environmental tax τ_a^s are not included in the objective function, because payment of a toll or tax implies only a transfer of money from road users to the authority inside the system, and not a deadweight loss.

Let P^s be the probability of scenario *s*. The expected value $E[Z]$ of the total system disutility over the scenario set Ω is then given by

$$E[Z] = \sum_{s\in\Omega} P^s Z^s \,. \tag{25}$$

We now introduce a measure of risk known as semi-deviation as an indicator of solution robustness, following Ahmed (2006) and Ruszczynski and Shapiro (2003). This measure is defined as

$$\Theta[Z] = E\left[Z - E[Z]\right]_+ = \sum_{s\in\Omega} P^s \left| Z^s - \sum_{s\in\Omega} P^s Z^s \right|, \tag{26}$$

where $[a]_+$ denotes the absolute value of *a*, or $[a]_+ = |a|$.

The semi-deviation, but not the variance (or standard deviation), is introduced as a proxy for the risk measure because variance is a symmetric statistic and gives equal weight to deviations above and below the mean without addressing the risks associated with extreme outcomes (List et al. 2003; Yin, Madanat, and Lu 2009), and because mean-variance does not preserve the convexity of the objective function (Ahmed 2006). The semi-deviation risk measure defined in Equation (26) overcomes both of these shortcomings.

We formulate a mean-risk robust toll design model as follows.

$$\min_{\mathbf{x}} \quad \varphi\big(\mathbf{x}, \mathbf{v}(\mathbf{x}), \mathbf{d}(\mathbf{x})\big) = E[Z] + \lambda \times \Theta[Z] = \sum_{s\in\Omega} P^s Z^s + \lambda \sum_{s\in\Omega} P^s \left| Z^s - \sum_{s\in\Omega} P^s Z^s \right|, \tag{27}$$

subject to

$$x_a^{\min} \le x_a \le x_a^{\max}, \quad \forall a \in \bar{A}, \tag{28}$$

where $\mathbf{v}(\mathbf{x})$ and $\mathbf{d}(\mathbf{x})$ can be calculated by solving the following scenario-based capacitated UE problem.

$$\min_{\mathbf{v}} \sum_{a\in A\setminus\bar{A}} \int_0^{v_a^s} \hat{t}_a^s(\omega)d\omega + \sum_{a\in\bar{A}} \int_0^{v_a^s} \left(\hat{t}_a^s(\omega) + \frac{x_a}{\beta} \right) d\omega, \tag{29}$$

subject to

$$\sum_{r\in R_w} f_{rw}^s = q_w^s, \quad \forall w \in W, s \in \Omega, \tag{30}$$

$$v_a^s = \sum_{w\in W} \sum_{r\in R_w} f_{rw}^s \delta_{ar}, \quad \forall a \in A, s \in \Omega, \tag{31}$$

$$v_a^s \le C_a, \quad \forall a \in A, s \in \Omega, \tag{32}$$

$$v_a^s \le \bar{E}_a, \quad \forall a \in A, s \in \Omega, \text{ and} \tag{33}$$

$$f_{rw}^s \ge 0, \quad \forall r \in R_w, w \in W, s \in \Omega. \tag{34}$$

In this model, $x_a^{\min}$ and $x_a^{\max}$ are the lower and upper bounds, respectively, of the toll charges on link a. λ is a non-negative weighting factor that reflects the significance of the risk measure for the objective function, and models the tradeoff between the average level of system performance (i.e., the expected total system disutility) and the robustness of the system (i.e., the semi-deviation risk measure). The larger the value of λ, the greater the robustness of the system, and vice versa. When λ equals zero, the mean-risk robust toll design model is reduced to a traditional expectation model. The expectation model is thus a special case of the robustness model.

We now introduce an additional variable, y^s, into the objective function (27) to eliminate the absolute value operation. This gives the following proposition.

Proposition 3. *The robust toll design model (27)-(34) can be rewritten as*

$$\min_{\mathbf{x},\mathbf{y}} \ \varphi\left(\mathbf{x}, \mathbf{v}(\mathbf{x}), \mathbf{d}(\mathbf{x}), \mathbf{y}\right) = \sum_{s\in\Omega} P^s Z^s + \lambda \sum_{s\in\Omega} P^s \left(Z^s - \sum_{s\in\Omega} P^s Z^s + 2y^s \right), \tag{35}$$

subject to

$$Z^s - \sum_{s\in\Omega} P^s Z^s + y^s \ge 0, \quad \forall s \in \Omega, \tag{36}$$

$$y^s \ge 0, \quad \forall s \in \Omega, \textit{ and} \tag{37}$$

$$x_a^{\min} \le x_a \le x_a^{\max}, \quad \forall a \in \bar{A}, \tag{38}$$

where $\mathbf{v}(\mathbf{x})$ *and* $\mathbf{d}(\mathbf{x})$ *can be given by the scenario-based capacitated UE problem (29)-(34).*

Proof. We consider two cases. In Case (i), if $Z^s - \sum_{s\in\Omega} P^s Z^s \geq 0$, then $y^s = 0$ and thus $\varphi = \sum_{s\in\Omega} P^s Z^s + \lambda \sum_{s\in\Omega} P^s \left(Z^s - \sum_{s\in\Omega} P^s Z^s \right)$. In Case (ii), if $Z^s - \sum_{s\in\Omega} P^s Z^s < 0$, then $y^s = \sum_{s\in\Omega} P^s Z^s - Z^s$ and thus $\varphi = \sum_{s\in\Omega} P^s Z^s + \lambda \sum_{s\in\Omega} P^s \left(\sum_{s\in\Omega} P^s Z^s - Z^s \right)$. This implies that the formulation (35)-(38) is identical to the original model (27)-(34).

The next section shows that this change in the model formulation is useful for developing a heuristic algorithm for solving the robust toll design model.

4. Solution procedure

4.1. Sample average approximation (SAA) based solution scheme

The idea behind the SAA scheme is to approximate a robust optimization problem using the sample average estimate derived from a random sample. The original robust optimization problem is then transformed into a deterministic equivalent that can be solved by the solution techniques for deterministic problems. The process is repeated with different samples to obtain candidate solutions, along with statistical estimates of their optimality gaps.

In the SAA-based solution scheme, we generate a sample of size *N*, or $\{s_1, s_2, \cdots, s_N\}$, for each OD demand by using Monte Carlo simulation. When generating multivariate, correlated random variables, the Monte Carlo simulation technique entails an excessively high computational effort (Daganzo, Bouthelier, and Sheffi 1977), and other approaches, such as Clark's approximation (Clark 1961) or the multivariate random variate generation procedure proposed in Chang, Tung, and Yang (1994), can be adopted as an alternative. The robust toll design model (35)-(38) can then be approximated by the following SAA problem.

$$\min_{\mathbf{x},\mathbf{y}} \ \varphi_N\left(\mathbf{x}, \mathbf{v}(\mathbf{x}), \mathbf{d}(\mathbf{x}), \mathbf{y}\right) = \frac{1}{N}\sum_{i=1}^{N} Z^{s_i} + \frac{\lambda}{N}\sum_{i=1}^{N}\left(Z^{s_i} - \frac{1}{N}\sum_{i=1}^{N} Z^{s_i} + 2y^{s_i} \right), \tag{39}$$

subject to

$$Z^{s_i} - \frac{1}{N}\sum_{i=1}^{N} Z^{s_i} + y^{s_i} \geq 0, \tag{40}$$

$$y^{s_i} \geq 0, \text{ and} \tag{41}$$

$$x_a^{\min} \leq x_a \leq x_a^{\max}, \quad \forall a \in \bar{A}, \tag{42}$$

where $\mathbf{v}(\mathbf{x})$ and $\mathbf{d}(\mathbf{x})$ can be determined by (29)-(34) for each scenario s_i.

The SAA problem in Equations (39)-(42) is a non-linear deterministic mathematical programming problem with a scenario-based capacitated UE problem forming the constraints, which can be solved using a sensitivity analysis-based algorithm. To implement the algorithm, we formulate the linear approximation of Z^{s_i} using its gradient information, as follows.

$$Z^{s_i} \approx Z^{s_i}(\mathbf{x}^*) + \nabla_{\mathbf{x}} Z^{s_i}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*), \tag{43}$$

where

$$\frac{\partial Z^{s_i}}{\partial x_a} = \beta \sum_{b \in A}\left[\frac{\partial v_b^{s_i}}{\partial x_a}\left(\hat{t}_b^{s_i} + d_b^{s_i}\right) + v_b^{s_i}\left(\frac{\partial t_b^{s_i}}{\partial v_b^{s_i}}\frac{\partial v_b^{s_i}}{\partial x_a} + \frac{\partial d_b^{s_i}}{\partial x_a}\right)\right] + \gamma \sum_{b \in A}\left[\frac{\partial v_b^{s_i}}{\partial x_a}\left(e_b^{s_i} + v_b^{s_i}\frac{\partial e_b^{s_i}}{\partial v_b^{s_i}}\right)\right], \tag{44}$$

where

$$\frac{\partial e_a^{s_i}}{\partial v_a^{s_i}} = 9.1913 \cdot \frac{\partial t_a^{s_i}}{\partial v_a^{s_i}} \cdot \left(1 - 0.01023\frac{l_a}{t_a^{s_i}(v_a^{s_i})}\right) \cdot \exp\left(0.01023\frac{l_a}{t_a^{s_i}(v_a^{s_i})}\right), \text{ and} \tag{45}$$

$$\frac{\partial t_a^{s_i}}{\partial v_a^{s_i}} = \frac{n\alpha l_a}{k_a^0 C_a}\left(\frac{v_a^{s_i}}{C_a}\right)^{n-1}, \tag{46}$$

where the derivatives $\partial v_b^{s_i} / \partial x_a$ and $\partial d_b^{s_i} / \partial x_a$ can be obtained using the sensitivity analysis method for network equilibrium problems originally proposed by Tobin and Friesz (1988) and later extended and applied by Yang and Yagar (1995), Yang and Lam (1996), and Yang and Bell (1997).

By substituting Equation (43) into Equations (39) and (40), SAA problem (39)-(42) then becomes a linear programming problem that can be solved using the well-known simplex method. The SAA-based solution algorithm is described as follows.

Step 0. *Initialization*. Generate a sample of size *N* for each OD demand, that is, *N* demand scenarios $\{s_1, s_2, \cdots, s_N\}$, and choose an initial toll solution $\mathbf{x}^{(0)} = \{x_a^{(0)}\}$. Set the iteration counter *j* equal to 0.

Step 1. *Solving of the second-stage problem.* For each scenario s_i, run the following steps.

Step 1.1. Solve the scenario-based capacitated UE problem (29)-(34) for a given $\mathbf{x}^{(j)}$, and obtain $\{v_a^{s_i(j)}\}$ and $\{d_a^{s_i(j)}\}$.

Step 1.2. Conduct a sensitivity analysis to calculate the derivatives $\partial v_b^{s_i(j)} \big/ \partial x_a$ and $\partial d_b^{s_i(j)} \big/ \partial x_a$.

Step 1.3. Calculate $\nabla_{\mathbf{x}} Z^{s_i(j)}$ and $Z^{s_i(j)}$ according to Equations (43)-(46).

Step 2. *Solving of the first-stage problem.* Formulate the local linear approximation of the SAA problem (39)-(42), and solve the resultant linear programming problem to obtain the auxiliary toll solution $\mathbf{h}^{(j)}$.

Step 3. *Updating*. Update the toll pattern in terms of $\mathbf{x}^{(j+1)} = \mathbf{x}^{(j)} + \left(\mathbf{h}^{(j)} - \mathbf{x}^{(j)}\right) \big/ \left(j+1\right)$.

Step 4. *Convergence check.* If $\max(|x_a^{(j+1)} - x_a^{(j)}|, \forall a \in \bar{A}) \le \varepsilon$, where ε is a pre-specified error tolerance, then stop. Otherwise, set $j = j + 1$ and go to Step 1.

It should be pointed out that different demand scenarios lead to different solutions. Intuitively, the larger the number of samples, the more robust the solution. However, the number of variables and constraints in the SAA problem (39)-(42) increase linearly with the sample size, leading to a rapid rise in the computational complexity of the model. Fortunately, however, Mulvey, Vanderbei, and Zenios (1995) and Laguna (1998) showed that a relatively small sample can produce a near-optimal solution. Kleywegt, Shapiro, and Homem-De-Mello (2001) showed that as the sample size increases, the solution to the SAA problem (39)-(42) converges exponentially quickly with a probability of 1 to the optimal solution of the original problem (27)-(34). In practice, the selection of the sample size for the SAA problem is a tradeoff between the quality or accuracy of the solution and the computational effort required to solve the problem.

4.2. Evaluation of the solution quality

The optimal solution obtained using this sampling technique cannot guarantee the optimality of the original problem. However, taking the optimal solutions from several sample sets provides a statistical inference of the confidence interval of the actual optimal solution.

Let $\mathbf{x}_N^*$ and φ_N^* be the optimal toll solution and optimal objective value, respectively, of an approximate problem with sample size *N*. In general, $\mathbf{x}_N^*$ and φ_N^* vary with the sample size *N*. Let $\mathbf{x}^*$ and φ^* be the optimal solution and optimal objective value, respectively, of the original problem. Clearly, the following relationship holds.

$$\varphi^* \le \varphi_N^* . \tag{47}$$

This means that φ_N^* is the upper bound of the optimal solution of the original problem.

Note that $\mathbf{x}_N^*$ is the optimal solution of the approximate problem, which gives

$$\varphi_N^* = \varphi_N(\mathbf{x}_N^*) \le \varphi_N(\mathbf{x}^*) . \tag{48}$$

Taking the expectations on both sides of Equation (48) yields

$$E\left[\varphi_N^*\right] \le E\left[\varphi_N(\mathbf{x}^*)\right]. \tag{49}$$

As the SAA problem is an unbiased estimator of the original problem, we have

$$E\left[\varphi_N^*\right] \le E\left[\varphi_N(\mathbf{x}^*)\right] = \varphi^* . \tag{50}$$

Equation (50) indicates that the expected value $E\left[\varphi_N^*\right]$ of φ_N^* is the lower bound of the optimal objective value φ^* of the original problem. In the following, the statistical lower and upper bound estimates of the true objective value are discussed.

4.2.1. Estimate of the lower bound

The expectation $E\left[\varphi_N^*\right]$ can be estimated by generating M_l independent sample groups, each with N_l samples. Solving the corresponding SAA problem (39)-(42) for each sample group *m* gives the optimal solution $\mathbf{x}_{N_l}^{m*}$ and the optimal objective value $\varphi_{N_l}^{m*}$. Following Equation (50), the lower bound of the original objective function (27) is then given by

$$\bar{\varphi}_l = E\left[\varphi^*_{N_l}\right] = \frac{1}{M_l}\sum_{m=1}^{M_l}\varphi^{m*}_{N_l}. \tag{51}$$

According to the central limit theorem, the distribution of the lower bound estimate of the original problem converges to a normal distribution $N\left(\mu_l, \sigma_l^2\right)$. The mean $\mu_l = E\left[\varphi^*_{N_l}\right]$ can be approximated by a sample mean $\bar{\varphi}_l$ (shown in Equation (51)), and the variance $\sigma_l^2 = \mathrm{var}\left[\varphi^*_{N_l}\right]$ can be approximated by a sample variance represented as

$$\sigma_l^2 = \frac{1}{\left(M_l - 1\right)M_l}\sum_{m=1}^{M_l}\left(\varphi^{m*}_{N_l} - \bar{\varphi}_l\right)^2. \tag{52}$$

The $\left(1-\theta\right)$ confidence interval for this lower bound is

$$\left[\bar{\varphi}_l - z_{\theta/2}\sigma_l, \bar{\varphi}_l + z_{\theta/2}\sigma_l\right], \tag{53}$$

where $z_{\theta/2}$ satisfies that the probability $\Pr\left\{-z_{\theta/2} \le N\left(0,1\right) \le z_{\theta/2}\right\} = 1-\theta$, where $N\left(0,1\right)$ denotes the probability density function of the standard normal distribution. For instance, the 95% confidence interval of the lower bound implies that $\theta = 0.05$ and $z_{\theta/2} = 1.96$.

4.2.2. Estimate of the upper bound

We select a feasible solution $\bar{\mathbf{x}}^*_N$ from the set of sample solutions $\{\mathbf{x}^{m*}_{N_l}\}, m = 1, 2, \cdots, M_l$, and generate M_u independent sample groups, each with N_u samples. The upper bound of the original problem can then be estimated by

$$\bar{\varphi}_u(\bar{\mathbf{x}}^*_N) = \frac{1}{M_u}\sum_{m=1}^{M_u}\varphi^{m*}_{N_u}(\bar{\mathbf{x}}^*_N). \tag{54}$$

According to the central limit theorem, the distribution of the upper bound estimate converges to a normal distribution $N\left(\mu_u, \sigma_u^2\right)$, where the mean $\mu_u = E\left[\varphi^*_{N_u}\right]$ can be approximated by a sample mean $\bar{\varphi}_u(\bar{\mathbf{x}}^*_N)$ (see Equation (54)) and the variance $\sigma_u^2 = \mathrm{var}\left[\varphi^*_{N_u}\right]$ can be estimated by a sample variance given by

$$\sigma_u^2 = \frac{1}{(M_u - 1)M_u} \sum_{m=1}^{M_u} \left(\varphi_{N_u}^{m*}(\bar{\mathbf{x}}_N^*) - \bar{\varphi}_u(\bar{\mathbf{x}}_N^*) \right)^2 . \quad (55)$$

Hence, the $(1-\theta)$ confidence interval for the upper bound is

$$\left[\bar{\varphi}_u(\bar{\mathbf{x}}_N^*) - z_{\theta/2}\sigma_u, \bar{\varphi}_u(\bar{\mathbf{x}}_N^*) + z_{\theta/2}\sigma_u \right], \quad (56)$$

where $z_{\theta/2}$ satisfies that the probability $\Pr\left\{-z_{\theta/2} \le N(0,1) \le z_{\theta/2}\right\} = 1-\theta$.

4.2.3. Optimality gap estimate

On the basis of the upper and lower bound estimates, we can compute the estimate for the optimality gap by

$$\bar{\varphi}_{\text{gap}} = \bar{\varphi}_u(\bar{\mathbf{x}}_N^*) - \bar{\varphi}_l . \quad (57)$$

The variance for the gap estimate is then given by

$$\sigma_{\text{gap}}^2 = \sigma_l^2 + \sigma_u^2 . \quad (58)$$

5. Numerical studies

To facilitate the presentation of the essential ideas and contributions of this study, in this section two test scenarios are used to illustrate the application of the proposed model and solution algorithm. This first scenario shows the properties of the proposed model and establishes the importance of the incorporation of risk and environmental taxation in road toll design. The second scenario shows the efficiency or performance of the proposed solution algorithm and investigates the effects of OD demand variation on the model solution and the computation time of the solution algorithm. The proposed solution algorithm was coded in programming language C and run on a personal computer with an Intel(R) Core(TM) I5-520M processor, 2.40 GHz, and 4GB of RAM. The stopping tolerance ε in both examples is set to 0.0001.

5.1. Scenario 1

The example network for Scenario 1, as shown in Figure 3, consists of four nodes, six links, and one OD pair (1-4). In Figure 3, node 3 (which is shaded) represents a signal-controlled

intersection. It is assumed that the proportion of green time at this intersection is 50% and the constant signal cycle time Γ is one minute. The free-flow speed k_a^0 of vehicles is 48 km/h on each link. The traveler's time value β is US$20/h, and the damage value γ per unit of pollutant emissions is US$2494/ton. The maximum permissible emission for each link is assumed to be 5,000 grams per hour. The parameters α and n in the link travel time function (6) are 0.15 and 4.0, respectively. The lengths and saturation capacities of all links are given in Table 2. The OD demand is assumed to follow a truncated normal distribution with $4200 \le N\left(5000,(300)^2\right) \le 5800$.

We first illustrate the relationship between the physical and environmental capacities, which is stated in Proposition 1. Table 3 shows the resultant critical length and physical and environmental capacities for each link. It can be seen that for links 1 to 5, the physical capacities are less than the associated environmental capacities, or $C_a < \bar{E}_a,\ \forall a = 1,...,5$. As a result, an environmental tax is not incurred on these five links. However, on link 6, the physical capacity exceeds the environmental capacity, with $C_6 > \bar{E}_6$ (shown in bold in the last row of Table 3), which indicates that the environmental capacity constraint of link 6 will become binding earlier than its physical capacity constraint. Accordingly, following Proposition 2, an extra environmental tax will be required on link 6.

It should be noted that the environmental capacities of different links for this example network are different even for links with the same physical capacity and length. For instance, links 2 and 3 have the same length of 7.2km and the same physical capacity of 1500 veh/h. However, their environmental capacities and critical lengths (see Table 3) are different due to the existence of a signal-controlled intersection at node 3. This indicates that the potential effect of signal intersections on traffic pollutant emissions should not be neglected. The incorporation of this effect into the optimization of signal settings is an interesting and important topic but is outside the scope of this paper, and is thus left to future research.

Figure 4 plots the expected total system disutility against its semi-deviation risk measure in the SAA problem (39)-(42) for different values of λ, which measures the risk-averse degree of the toll design scheme. The sample size adopted in the implementation of the SAA scheme is 500. Figure 4 shows that as λ increases from 0 (i.e., the expectation model) to 100 (i.e., a

risk-averse model), the expected total system disutility increases from \$20,482 to \$21,742 per hour, whereas the corresponding semi-deviation risk measure decreases from \$654 to \$511 per hour. This means that a larger value of λ leads to more robust or reliable system performance, and vice versa. Ignoring the risk term (i.e., the traditional expectation model) can lead to the underestimation of the expected total system disutility and the overestimation of its variation. Consequently, in a robust toll design, a tradeoff should be made between the expected system performance and its risk.

We now compare the results generated by different pricing schemes when λ is fixed at 10. In Scheme I, links 5 and 6 are subject to a congestion toll ranging between \$0.00 and \$2.00 and an environmental tax, respectively. In Scheme II, only link 5 is subject to a congestion toll. In Scheme III, no link toll is required in the network (i.e., there is no intervention). To estimate the lower and upper bounds of the objective function in the original robust design model (27)-(34) for different pricing schemes, 10 sample groups with 500 samples in each group are used, that is, $M_l = M_u = 10$ and $N_l = N_u = 500$. Table 4 shows that among the schemes, Scheme III generates the worst network performance (the largest lower/upper bounds of the objective value) and Scheme I the best network performance. This suggests that a scheme that combines a congestion toll and an environmental tax is the most productive and efficient in terms of road network performance, and can serve as a useful tool for managing traffic congestion and vehicular emissions.

5.2. Scenario 2

The network for Scenario 2 is shown in Figure 5, which is adopted from Nguyen and Dupuis (1984) and consists of 13 nodes, 19 links, and 4 OD pairs (i.e., 1-2, 1-3, 4-2, and 4-3). Nodes 5, 6, 8, 9, 10, and 11 in the network represent signal-controlled intersections. Links 13 and 15 are two bottleneck links that are subject to congestion tolls ranging from \$0.00 to \$2.00. The OD demands are assumed to follow the following independent (truncated) normal distributions.

$$\begin{cases} 1800 \le N\left(2500, (300)^2\right) \le 3200, & \text{for OD pairs (1, 2) and (4, 3),} \\ 1000 \le N\left(1500, (200)^2\right) \le 2000, & \text{for OD pairs (1, 3) and (4, 2).} \end{cases} \tag{59}$$

The parameters of the link travel time functions are given in Table 5. Again, the proportion of

green time and the constant signal cycle time at each signal-controlled intersection are assumed to be 50% and one minute, respectively. The other parameters are $k_a^0 = 48$ km/h, $\beta = 20$ US$/h, $\gamma = 2494$ US$/ton, $E_a = 5000$ g/h, $\alpha = 0.15$, and $n = 4.0$.

Table 6 shows the resultant critical length and physical and environmental capacities for each link. It can be seen that for links 4 and 18, the environmental capacities are less than their associated physical capacities, or $\bar{E}_4 < C_4$ and $\bar{E}_{18} < C_{18}$, as shown in bold in Table 6. As a result, an additional environmental tax is required on links 4 and 18 according to Proposition 2.

Figure 6 illustrates the changes in the value of objective function $\varphi_N(\cdot)$ in Equation (39) and in the CPU time required to compute the model solution for numbers of demand scenarios (the sample size for each OD demand) varying from 5 to 800. It can be seen that the value of $\varphi_N(\cdot)$ varies sharply when the number of demand scenarios increases from 5 to 400. However, after 500 scenarios, the value of $\varphi_N(\cdot)$ stabilizes. This means that a sufficiently large number of samples results in a good approximation of the original or true problem with a continuous OD demand distribution. Figure 6 also shows that the computational time (in CPU seconds) increases with the number of scenarios. Based on a tradeoff between the quality or accuracy of the solution and the computational effort, the numerical results presented below are generated with 500 samples.

We now look at the effects of the OD demand variation by checking two demand variation levels: the base case (the demand variation shown in Equation (59)) and twice the base case. Table 7 shows the effects of OD demand variation on the model solution and the computational time. When the level of demand variation for each OD pair doubles, the value of the objective function $\varphi_N(\cdot)$ increases by \$7812 per hour (from \$79406 to \$87218 per hour). This means that a large demand fluctuation can diminish the performance of the transportation system. Further, as the OD demand variation increases, the CPU time required to solve the proposed robust design model increases by 283 seconds (from 1,383 to 1,666 seconds).

Figure 7 plots the histograms of the total system travel disutility and the total amount of

emissions for the aforementioned two OD demand variation levels under their associated optimal tolls, as shown in Table 7. It can be observed that a larger OD demand variation leads to a broader spread in both the total system travel disutility and the total amount of emissions. This further illustrates that, as the OD demand variation increases, the performance of the transportation system diminishes.

6. Conclusion and further studies

In this study, a new two-stage robust optimization model is proposed to investigate road toll design problems for policy assessment and strategic planning purposes. In the proposed model, the effects of demand uncertainty are explicitly considered together with congestion and environmental externalities caused by vehicular traffic on the roads. Based on analyses of physical and environmental capacity constraints, the boundary conditions under which a road user on a link should pay either a congestion toll or an extra environmental tax are determined. A heuristic solution algorithm is proposed and its solution quality evaluated. Two example networks are presented to illustrate the properties of the proposed model and solution algorithm. The proposed model is shown to be a useful tool for studying congestion and environmental pricing in urban road networks with uncertainty, and can be used to assess the effects of various travel demand management measures and vehicular emission policies at a strategic level.

Although the numerical results for small networks easily illustrate the essential merits of the proposed model, we recognize that case studies of large and realistic networks are necessary to further validate the findings of the numerical examples and the performance of the proposed model. In this study, attention has mainly focused on demand uncertainty. However, it is of great importance to extend this study to incorporate the effects of supply uncertainty on travel time variability to further improve the robustness of the model. Future studies could also focus on extending the proposed model to consider pollutant emissions from different modes of transportation in urban transport systems, to optimize traffic signal settings at road intersections, and to incorporate the risk-taking behavior of travelers in view of network uncertainties.

Acknowledgements

The work that is described in this paper was jointly supported by grants from the Research Committee of the Hong Kong Polytechnic University (Project nos. G-YX1V and 1-ZV4T), the National Natural Science Foundation of China (70971045), the Research Foundation for the Author of National Excellent Doctoral Dissertation (China) (200963), the Program for New Century Excellent Talents in University (China) (NCET-10-0385), and the University Research Committee and HKU SPACE Research Fund of the University of Hong Kong (201007176059).

References

Ahmed S. 2006. Convexity and decomposition of mean-risk stochastic programs. Mathematical Programming 106(3): 433-446.

Bell MGH. 1995. Stochastic user equilibrium assignment in networks with queues. Transportation Research Part B 29(2): 125-137.

Benedek CM, Rilett LR. 1998. Equitable traffic assignment with environmental cost functions. Journal of Transportation Engineering – ASCE 124(1): 16-22.

Black WR, Sato N. 2007. From global warming to sustainable transport 1989-2006. International Journal of Sustainable Transportation 1(2): 73-89.

Boyles SD, Kockelman KM, Waller ST. 2010. Congestion pricing under operational, supply-side uncertainty. Transportation Research Part C 18(4): 519-535.

Chang CH, Tung YK, Yang JC. 1994. Monte Carlo simulation for correlated variables with marginal distributions. Journal of Hydraulic Engineering 120(3): 313-331.

Chen A, Yang H, Lo HK, Tang WH. 2002. Capacity reliability of a road network: An assessment methodology and numerical results. Transportation Research Part B 36: 225-252.

Chen A, Kim J, Lee S, Kim Y. 2010. Stochastic multi-objective models for network design problem. Expert Systems with Application 37: 1608-1619.

Chen A, Subprasom K, Ji Z. 2003. Mean-variance model for the build-operate-transfer scheme under demand uncertainty. Transportation Research Record 1857: 93-101.

Chen A, Zhou Z, Ryu S. 2011. Modeling physical and environmental side constraints in traffic equilibrium problem. International Journal of Sustainable Transportation 5(3): 172-197.

Chootinan P, Wong SC, Chen A. 2005. A reliability-based network design problem. Journal of

Advanced Transportation 39(3): 247-270.

Clark A, Sumalee A, Shepherd S, Connors R. 2009. On the existence and uniqueness of first best tolls in networks with multiple user classes and elastic demand. Transportmetrica 5(2): 141-157.

Clark CE. 1961. The greatest of a finite set of random variables. Operations Research 9(2): 145-162.

Daganzo CF, Bouthelier F, Sheffi Y. 1977. Multinomial Probit and qualitative choice: A computationally efficient algorithm. Transportation Science 11(4): 338-358.

Epperson JF. 2007. An introduction to numerical methods and analysis. John Wiley & Sons, NJ, USA.

Ferrari P. 1995. Road pricing and network equilibrium. Transportation Research Part B 29(5): 357-372.

Ferrari P. 1997. Capacity constraints in urban transport networks. Transportation Research Part B 31(4): 291-301.

Gardner LM, Unnikrishnan A, Waller ST. 2008. Robust pricing for transportation networks with demand uncertainty. Transportation Research Record 2085: 21-30.

Hau TD. 2005. Economics fundamentals of road pricing: A diagrammatic analysis, Part I – Fundamentals. Transportmetrica 1(2): 81-117.

Hau TD. 2006. Congestion charging mechanisms for roads, Part II – Case studies. Transportmetrica 2(2): 117-152.

Hearn DW, Ramana MV. 1998. Solving congestion toll pricing models. In: Marcotte P, Nguyen S. (eds), Equilibrium and Advanced Transportation Modeling. Kluwer Academic Publishers, Norwell.

IPCC (Intergovernmental Panel on Climate Change). 2007. Fourth Assessment Report: Climate Change 2007. Available at: http://www.ipcc.ch/publications_and_data/publications_and_data_reports.htm.

Jakkula N, Asakura Y. 2009. Accuracy of optimum road pricing considering local emissions of road traffic network control in transportation systems. Paper presented at the 12th IFAC (International Federation of Automatic Control) Symposium on Control in Transportation Systems, September 2-4, 2009, Redondo Beach, California, USA.

Johansson O. 1997. Optimal road-pricing: Simultaneous treatment of time losses, increased fuel consumption, and emissions. Transportation Research Part D 2(2): 77-87.

Johansson O. 2006. Optimal environmental road pricing. Economic Letters 90(2): 225-229.

Kim J, Kurauchi F, Uno N. 2011. Analysis of variation in demand and performance of urban

expressways using dynamic path flow estimation. Transportmetrica 7(1): 63-84.

Kleywegt AJ, Shapiro A, Homem-De-Mello T. 2001. The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 12(2): 479-502.

Lam WHK, Shao H, Sumalee A. 2008. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. Transportation Research Part B 42(10), 890-910.

Laguna M. 1998. Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. Management Science 44(11): 101-110.

Larsson T, Patriksson M. 1995. An augmented Lagrangian dual algorithm for link capacity side constrained traffic assignment problems. Transportation Research Part B 29(6): 433-455.

Li H, Bliemer MCJ, Bovy PHL. 2008. Network reliability-based optimal toll design. Journal of Advanced Transportation 42(3): 311-332.

Li ZC, Huang HJ, Lam WHK, Wong SC. 2007. A model for evaluation of transport policies in multimodal networks with road and parking capacity constraints. Journal of Mathematical Modelling and Algorithms 6(2): 239-257.

Li ZC, Lam WHK, Wong SC, Huang HJ, Zhu DL. 2008. Reliability evaluation for stochastic and time-dependent networks with multiple parking facilities. Networks and Spatial Economics 8(4): 355-381.

List GF, Wood B, Nozick LK, Turnquist MA, Jones DA, Kjeldgaard EA, Lawton CR. 2003. Robust optimization for fleet planning under uncertainty. Transportation Research Part E 39(3): 209-227.

Lo HK, Luo XW, Siu BWY. 2006. Degradable transport network: Travel time budget of travellers with heterogeneous risk aversion. Transportation Research Part B 40(9): 792-806.

Mulvey JM, Vanderbei RJ, Zenios SA. 1995. Robust optimization of large-scale systems. Operations Research 43(2): 264-281.

Nagurney A. 2000a. Sustainable Transportation Networks. Edward Elgar Publishers, Cheltenham, UK.

Nagurney A. 2000b. Congested urban transportation networks and emission paradoxes. Transportation Research Part D 5(2), 145-151.

Nagurney A, Qiang Q, Nagurney LS. 2010. Environmental impact assessment of transportation networks with degradable links in an era of climate change. International

Journal of Sustainable Transportation 4(3): 154-171.

Nguyen S, Dupuis C. 1984. An Efficient method for computing traffic equilibria in networks with asymmetric transportation costs. Transportation Science 18(2): 185-202.

Penic MA, Upchurch J. 1992. TRANSYT-7F: Enhancement for fuel consumption, pollution emissions, and user costs. Transportation Research Record 1360: 104-111.

Rilett LR, Benedek CM. 1994. Traffic assignment under environmental and equity objectives. Transportation Research Record 1443: 92-99.

Roth IF, Ambs LL. 2004. Incorporating externalities into a full cost approach to electric power generation life-cycle costing. Energy 29(12-15): 2125-2144.

Rouwendal J, Verhoef ET. 2006. Basic economic principles of road pricing: From theory to applications. Transport Policy 13(2): 106-114.

Ruszczynski A, Shapiro A. 2003. Stochastic Programming. Elsevier, Amsterdam.

Santos G. 2004. Road Pricing: Theory and Evidence. Elsevier, Amsterdam.

Sugawara S, Niemeier DA. 2002. How much can vehicle emissions be reduced? Transportation Research Record 1815: 29-37.

Sumalee A, Watling DP, Nakayama S. 2006. Reliable network design problem: Case with uncertain demand and total travel time reliability. Transportation Research Record 1964: 81-90.

Sumalee A, Xu W. 2011. First-best marginal cost toll for a traffic network with stochastic demand. Transportation Research Part B 45(1): 41-59.

Szeto WY, Li X, O'Mahony M. 2008. Simultaneous occurrence of Braess and emission paradoxes. Proceedings of the Sixth International Conference on Traffic and Transportation Studies, pp. 625-634.

Taylor MAP. 1996. Incorporating environmental planning decisions in transport planning: A modelling framework. In: Hayashi Y, Roy J. (eds), Transport, Land-use and the Environment. Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 337-358.

Tobin RL, Friesz TL. 1988. Sensitivity analysis for equilibrium network flows. Transportation Science 22(4): 242-250.

Tsekeris T, Voß S. 2009. Design and evaluation of road pricing: State-of-the-art and methodological advances. Netnomics 10(1): 5-52.

USEPA (United States Environmental Protection Agency). 1991. National Air Quality and Emissions Trends Report. Available at: http://nepis.epa.gov.

USEPA. 1992. Transportation and Air Quality Planning Guidelines. Available at: http://nepis.epa.gov.

Verhoef ET. 2002. Second-best congestion pricing in general static transportation networks with elastic demands. Regional Science and Urban Economics 32(3): 281-310.

Wallace CE, Courage KG, Hadi MA, Gan AG. 1998. TRANSYT-7F User's Guide. University of Florida, Gainesville.

Wong SC, Yang H. 1997. Reserve capacity of a signal-controlled road network. Transportation Research Part B 31(5): 397-402.

Yang H, Bell MGH. 1997. Traffic restraint, road pricing and network equilibrium. Transportation Research Part B 31(4): 303-314.

Yang H, Huang HJ. 2005. Mathematical and Economic Theory of Road Pricing. Elsevier, Oxford.

Yang H, Lam WHK. 1996. Optimal road tolls under conditions of queuing and congestion. Transportation Research Part A 30(5): 319-332.

Yang H, Yagar S. 1995. Traffic assignment and signal control in saturated road networks. Transportation Research Part A 29(2): 125-139.

Yin Y, Lawphongpanich S. 2006. Internalizing emission externality on road networks. Transportation Research Part D 11(4): 292-301.

Yin Y, Madanat SM, Lu XY. 2009. Robust improvement schemes for road networks under demand uncertainty. European Journal of Operational Research 198(2): 470-479.

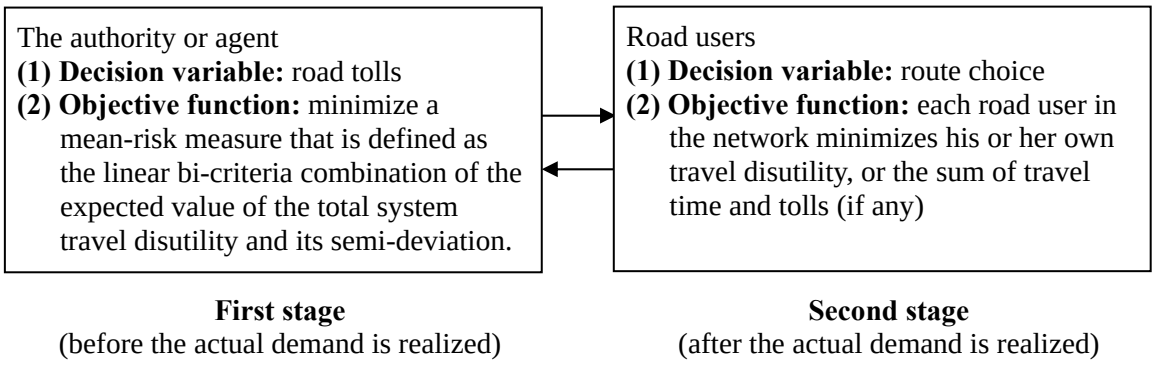


Figure 1. Framework of the two-stage decision model.

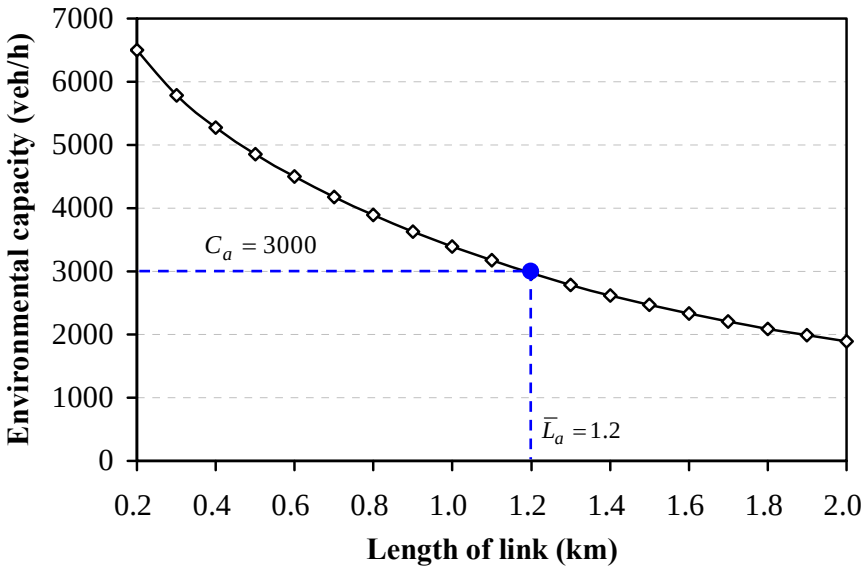


Figure 2. Environmental capacity versus length of link.

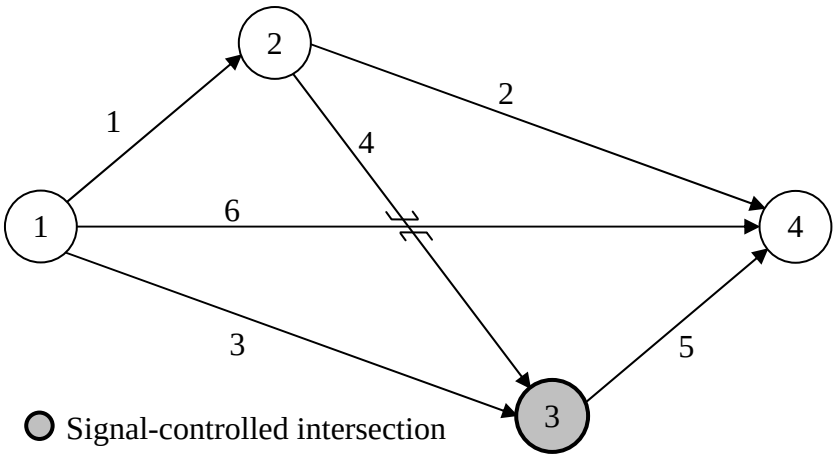


Figure 3. Example network for Scenario 1.

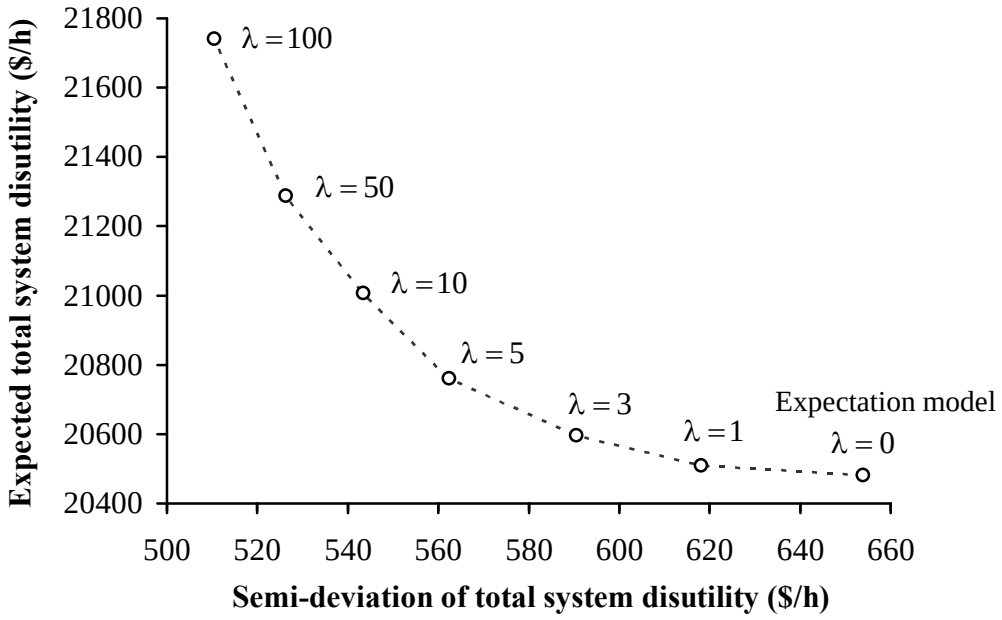


Figure 4. Expected total system disutility versus the semi-deviation risk measure in the SAA problem (39)-(42).

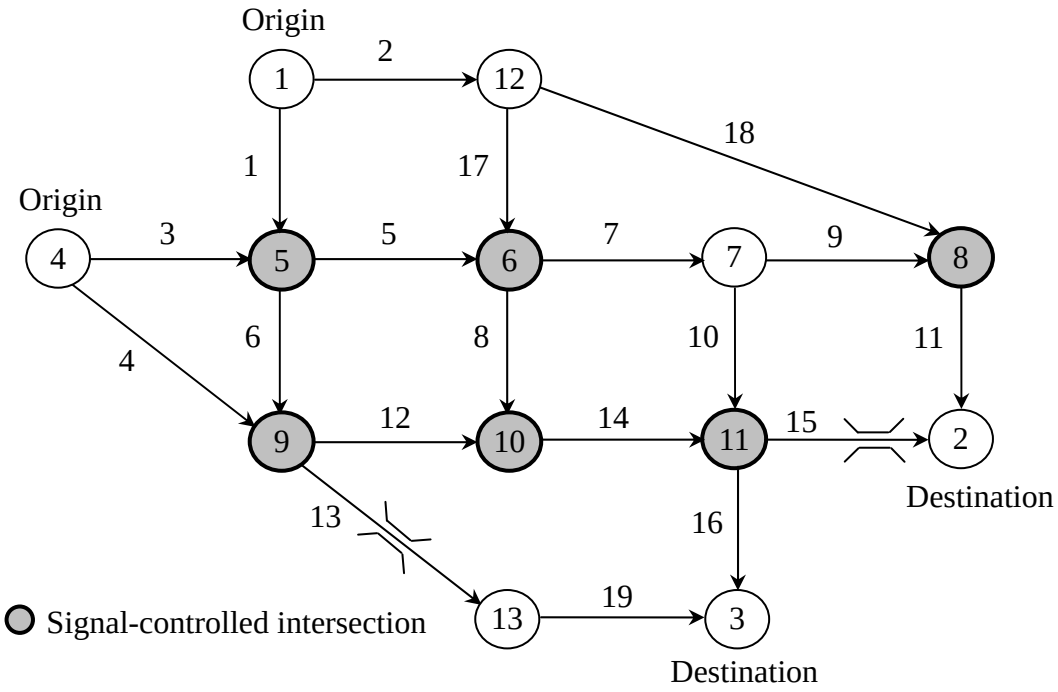


Figure 5. Nguyen and Dupuis's network for Scenario 2.

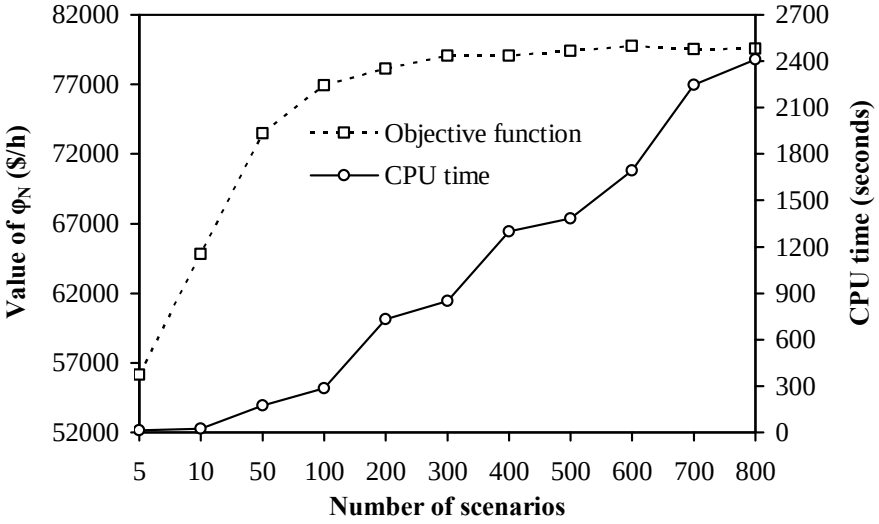
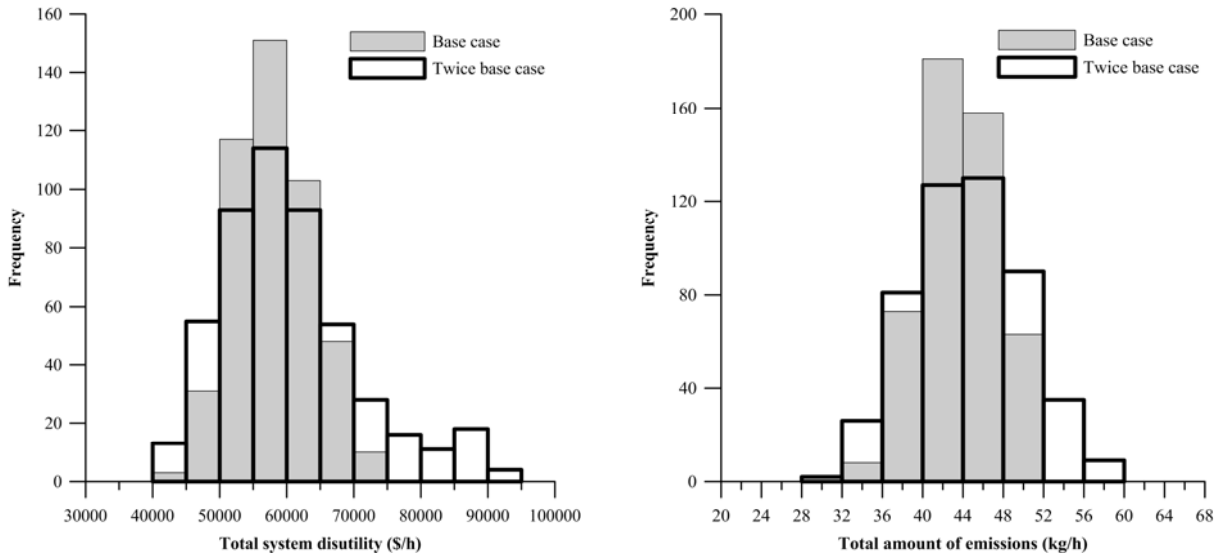


Figure 6. Value of objective function $\varphi_N(\cdot)$ in the SAA problem (39)-(42) and the CPU time of the proposed solution algorithm versus number of scenarios.



(a) **(b)**

Figure 7. Effects of OD demand variation on the distribution of (a) total system disutility and (b) total amount of emissions.

Table 1 Contributions to road toll design research

Type of road toll pricing	Deterministic model	Stochastic model
Congestion pricing	Ferrari (1995, 1997); Hearn and Ramana (1998); Yang and Lam (1996); Yang and Bell (1997); Yang and Huang (2005); Verhoef (2002); Clark et al. (2009)	Gardner, Unnikrishnan, and Waller (2008); Li, Bliemer, and Bovy (2008); Boyles, Kockelman, Waller (2010); Sumalee and Xu (2011)
Environmental pricing	Johansson (2006); Yin and Lawphongpanich (2006)	This paper
Simultaneous congestion and environmental pricing	Johansson (1997); Nagurney (2000a); Jakkula and Asakura (2009)	This paper

Table 2 Parameters of the link travel time functions in Scenario 1

Link No.	Link length l_a (km)	Saturation capacity $\bar{C}_a$ (veh/h)
1	4.8	3000
2	7.2	1500
3	7.2	3000
4	2.4	4000
5	2.4	3000
6	8.0	2000

Table 3 Critical link lengths and physical / environmental capacities for Scenario 1

Link No.	Critical length $\bar{L}_a$ (km)	Physical capacity C_a (veh/h)	Environmental capacity $\bar{E}_a$ (veh/h)
1	4.94	3000	3066
2	9.88	1500	1862
3	9.61	1500	1824
4	7.14	2000	3484
5	4.94	3000	4578
6	**7.41**	**2000**	**1882**

Note: The bolded row implies that the environmental capacity is less than the associated physical capacity.

Table 4 Comparison of solutions for different pricing schemes ($\lambda = 10$)

	Scheme I	Scheme II	Scheme III
Congestion toll x_5 ($)	1.06	0.90	—
Environmental tax τ_6 ($)	1.95	—	—
Lower bound $\bar{\varphi}_l$ ($/h)	25351	29189	33842
Std. dev. of lower bound σ_l ($/h)	73.4	187.2	398.3
95% conf. int. of lower bound ($/h)	[25207, 25495]	[28822, 29556]	[33062, 34623]
Upper bound $\bar{\varphi}_u$ ($/h)	27126	32992	34378
Std. dev. of upper bound σ_u ($/h)	41.3	158.6	180.8
95% conf. int. of upper bound ($/h)	[27045, 27207]	[32681, 33303]	[34024, 34732]
Gap $\bar{\varphi}_{\text{gap}}$ ($/h)	1775	3803	536
Std. dev. of gap σ_{gap} ($/h)	84.2	245.3	437.4

Note: std. dev. = standard deviation; conf. int. = confidence interval.

Table 5 Parameters of the link travel time functions in Scenario 2

Link No.	Link length l_a (km)	Saturation capacity $\bar{C}_a$ (veh/h)
1	3.6	5000
2	3.6	3000
3	3.6	6000
4	6.0	5000
5	3.6	6000
6	3.6	6000
7	3.6	3000
8	3.6	6000
9	3.6	3000
10	3.6	5000
11	3.6	3000
12	3.6	6000
13	5.4	2500
14	3.6	6000
15	3.6	2500
16	3.6	3000
17	3.6	3000
18	7.2	4000
19	3.6	3000

Table 6 Critical link lengths and physical / environmental capacities for Scenario 2

Link No.	Critical length $\bar{L}_a$ (km)	Physical capacity C_a (veh/h)	Environmental capacity $\bar{E}_a$ (veh/h)
1	5.66	2500	3324
2	4.94	3000	3724
3	4.67	3000	3574
4	**5.66**	**2500**	**2392**
5	4.67	3000	3574
6	4.67	3000	3574
7	4.94	3000	3724
8	4.67	3000	3574
9	9.61	1500	2521
10	5.66	2500	3324
11	4.94	3000	3724
12	4.67	3000	3574
13	5.93	2500	2679
14	4.67	3000	3574
15	5.93	2500	3434
16	4.94	3000	3724
17	9.61	1500	2521
18	**7.14**	**2000**	**1988**
19	4.94	3000	3724

Note: The bolded rows imply that the environmental capacity is less than the associated physical capacity.

Table 7 Effects of OD demand variation on the model solution and computational time ($\lambda = 10$)

Solution		Base level of demand variation	Twice the base level of demand variation
Congestion toll ($)	x_{13}	0.97	1.08
	x_{15}	0.46	0.43
Environmental tax ($)	τ_4	0.29	0.34
	τ_{18}	0.51	0.97
Value of objective function $\varphi_N(\cdot)$ ($/h)		**79406**	**87218**
CPU time (seconds)		**1383**	**1666**