

Liu, Qiumin; van den Berg, Vincent A. C.; Verhoef, Erik T.; Rui, Jian

Working Paper

Pricing in the stochastic bottleneck model with price-sensitive demand

Tinbergen Institute Discussion Paper, No. TI 2024-011/VIII

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Liu, Qiumin; van den Berg, Vincent A. C.; Verhoef, Erik T.; Rui, Jian (2024) : Pricing in the stochastic bottleneck model with price-sensitive demand, Tinbergen Institute Discussion Paper, No. TI 2024-011/VIII, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/298018>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2024-011/VIII
Tinbergen Institute Discussion Paper

Pricing in the Stochastic Bottleneck Model with Price-Sensitive Demand

*Qiumin Liu*¹

*Vincent A.C. van den Berg*²

*Erik T. Verhoef*³

*Rui Jian*⁴

¹ Beijing Transport Institute, VU Amsterdam and Beijing Jiaotong University

² VU Amsterdam and Tinbergen Institute

³ VU Amsterdam and Tinbergen Institute

⁴ Beijing Jiaotong University

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Pricing in the Stochastic Bottleneck Model with Price-Sensitive Demand

Qiumin Liu ^{a,b,c}, Vincent A.C. van den Berg ^{c,d,#}, Erik T. Verhoef ^{c,d}, Rui Jiang ^b

^a *Beijing Transport Institute, Beijing 100073, China*

^b *School of Systems Science, Beijing Jiaotong University, Beijing 100044, China*

^c *Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*

^d *Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands*

Corresponding author: v.a.c.vanden.berb@vu.nl

Abstract

We analyse time-varying tolling in the stochastic bottleneck model with price-sensitive demand and uncertain capacity. We find that price sensitivity and its interplay with uncertainty have important implications for the effects of tolling on travel costs, welfare and consumers. We evaluate three fully time-variant tolls and a step toll that have been proposed in previous literature. We also consider a uniform toll, which affects overall demand but not trip timing decisions.

The first fully time-variant toll is the “first-best” toll, which varies non-linearly over time and results in a departure rate that also varies over time. It raises the generalised price (i.e. the sum of travel cost and toll), thus lowering demand. These outcomes differ fundamentally from those found for first-best pricing in the deterministic bottleneck model. The second we call a “second-best” toll, as it aims to keep the departure rate constant over time. This is optimal without uncertainty, but it lowers welfare with uncertain capacity. Next, “third-best” tolling adds the further constraint that, besides a constant departure rate, the generalised price should stay the same as without tolling. It attains a lower welfare and higher expected travel cost than the second-best scheme, but a lower generalised price. All our tolling schemes—except for the third-best one—raise the price compared to the no-toll case.

In our numerical study, when there is less uncertainty, the second-best and third-best tolls achieve welfares closer to that of the first-best toll, and the three schemes are identical without uncertainty. As the degree of uncertainty falls, the uniform and single-step tolls attain welfare gains closer to that from first-best pricing. Also, when demand becomes more price-sensitive, the uniform and single-step tolls attain outcomes closer to those in the first-best outcome. Our ADL equilibrium step toll would lower the generalised price without uncertainty but raises it in our stochastic setting.

Keywords: stochastic bottleneck model; price-sensitive demand; time-varying toll; step toll; uncertainty

JEL codes: R41; R48; D62; D80

1. Introduction

Congestion is one of the greatest challenges for cities worldwide. It brings disutilities of different forms, including the pure loss of time, inconveniences from rescheduling in response to congestion, and additional inconveniences from unpredictability and uncertainty. Travel conditions may vary due to a combination of exogenous shocks, including weather conditions and traffic incidents, and endogenous demand responses to those shocks. Especially when those responses take the form of rescheduling—for example, by departing earlier to create time “buffers”—the analysis of congestion and policies to address it requires models that can deal with the dynamics of departure times, the impact of that on dynamic patterns of travel delays and the feedback of that upon behaviour, so as to obtain a consistent representation of stochastic dynamic equilibria that can be used for policy evaluation. That is what our paper aims to offer, studying policies to address dynamic congestion in a stochastic setting where travellers can respond in terms of departure time choice, but also travel choices more generally in the sense that we will consider price-sensitive demand.

Our analysis will employ the stochastic bottleneck model. We analyse different types of fully time-variant tolling, uniform tolling and step tolling; all with uncertain capacity and price-sensitive demand. As discussed below, there is a voluminous literature on the untolled equilibrium of the bottleneck model with uncertain capacity. Only a few papers consider time-variant tolling, and none also consider the price sensitivity of demand. As we will see, however, the interplay between uncertainty and price sensitivity has important effects and complicates the analysis. The effects of time-variant tolling will also change from those in the deterministic bottleneck model. Some papers have looked at uniform tolling, where the toll is constant (Lam, 2000; Jiang et al., 2021; Zhu et al., 2018; Yu et al., 2023a). Zhu et al. (2018) used a bottleneck model with uncertainty in the free-flow travel time that does not affect queuing. So, there is no interaction between uncertainty and the dynamics of congestion, and the model provides insights that are more similar to those in the deterministic bottleneck model, as the social optimum has no queuing. Conversely, in our model, the social optimum will have queuing when the uncertain capacity turns out to be very low. This seems more realistic, but it is also more difficult to analyse. We will provide a fuller literature review below.

Our methodological contribution is to study time-variant tolling in the stochastic bottleneck model with uncertain capacity and price-sensitive demand. We consider three fully time-variant tolling schemes, a step toll and a uniform toll. We optimise in two stages. The first, comparably to existing works with fixed demand, involves optimising the departure rate and the corresponding toll pattern for a given number of travellers, using optimal control theory. The second stage then optimises total demand, and therewith the time-invariant component or starting level of the toll, taking into account the effect on the first stage. This second stage, which is also relevant when the total number of car travellers can vary in response to policies, has not been considered before and will be shown to have important implications.

The three fully time-variant tolls that we consider follow from earlier studies with fixed demand. The first follows Lindsey (1994, 1999), and we call it “first-best” as it attains the overall social optimum. It leads to a departure rate that weakly increases over the morning, and it raises the generalised price (i.e. the sum of travel cost and toll) compared to the untolled equilibrium. It may also have a smaller reduction in travel cost and a welfare increase compared to under certainty, where first-best tolling leaves the price unchanged and halves the average travel cost. This has important implications for the political feasibility of tolling and its desirability versus alternative policies such as capacity expansion, information provision, travel credits and flexible working hours. Our results deviate from those for a

deterministic bottleneck in ways comparable to those under dynamic flow congestion, as in Chu (1999) and Mun (1999, 2003).

Long et al. (2022) proposed what we call a “second-best time-varying toll”, which is designed to achieve a departure rate that is constant over time. This simplifies the design of the tolling scheme for the government, and it would match what is optimal in the deterministic bottleneck model; but, as we will show, it is “second best” under uncertain capacity as it lowers welfare compared to first-best pricing.

Our third toll follows Xiao et al. (2015). We call it a “third-best time-variant toll” as it adds the further constraint (to the second-best setting) that the generalised price should remain the same as in the no-toll case. This raises the political feasibility of tolling, but as we will show, it leads to an even lower welfare and higher travel costs than the second-best scheme.

The analysis of uniform (time-invariant) and step tolling (where the toll changes only in discrete steps) is important. In reality, tolls are not fully time-variant; they are uniform—as in London—or at most have a few steps in them—as in Singapore and on some US pay-lanes. We will see that the combination of uncertainty and price-sensitive demand changes how these coarse tolls perform compared to the first-best one. Indeed, a uniform toll has no effect if demand is fixed: it cannot alter departure rates directly, it can only affect total demand. We use an ADL equilibrium step toll, which lowers the generalised price in the deterministic bottleneck model but raises it in our stochastic setting. Xiao et al. (2015) and Long et al. (2022) also considered single-step tolling in addition to their fully time-variant tolls. Jiang et al. (2022) studied single-step tolling but used fixed demand. Yu et al. (2023a) considered uniform tolling in conjunction with information provision. Jiang et al. (2021) and Zhu et al. (2018) analysed uniform tolling in a bottleneck model with an uncertain free-flow travel time.

Table 1 shows how our work relates to the closest literature and how we extend it, confirming that no previous study considered time-varying pricing in the bottleneck with uncertain capacity and price-sensitive demand. The remainder of the paper continues as follows. Section 2 overviews the broader literature, while the introduction focuses on the most related works. Section 3 gives the basic formulations for the no-toll equilibrium. Section 4 derives the social optimum of the time-varying toll and compares it to that of the deterministic model and the second- and third-best tolls. Sections 5 and 6 look at the uniform and single-step toll. Section 7 conducts a numerical study to illustrate the results and provides further insights. Section 8 concludes.

2. Extended literature review

Early works on uncertainty in the bottleneck model include Arnott et al. (1991, 1993b, 1996, 1999) and Daniel (1995). Arnott and co-authors were primarily interested in the effects of information provision and Daniel in competition among airlines. There is a very large literature on uncertainty in the bottleneck model. Small and Verhoef (2007), Small (2015) and Li et al. (2020), among others, provide extensive overviews.¹ But few papers look at congestion pricing,² and fewer still include price-sensitive demand; most look at the untolled equilibrium or information provision.

¹ There is also a large literature that considers static congestion (see Zhang et al. (2022) for a detailed review). In order to examine the interaction between information and pricing instruments, Verhoef et al. (1996) used the static model. They found that information and tolling are nearly perfectly complementary in the face of stochastic congestion. This has then been extended by also considering networks (Yang, 1999; Maher et al., 2005; Meng and Liu, 2011; Lindsey et al., 2014; Klein et al., 2018).

² Many papers have looked at alternative policies to reduce congestion and uncertainty. These include: information provision (e.g. Arnott et al., 1991, 1993b, 1996, 1999; Liu and Liu, 2018; Zhu et al., 2019; Yu et al., 2021; Han et al., 2021; Yu et al., 2023a), ride sharing (Long et al., 2018; Li et al., 2022; Liang et al., 2023), on-demand buses (Ma et al., 2023), tradable credits (Zhang et al., 2022), flexible working hours (Xiao et al., 2014b) and merging rules (e.g. Xiao et al., 2014a). Fosgerau (2010) considered the relation between the mean and variance of delays. Lindsey (2009) studied self-financing under random capacity and demand.

Table 1: Comparing our paper with the literature

Citation	Distribution of uncertainty	Bottleneck, uncertain capacity	Bottleneck, uncertain free flow time	Price-sensitive demand	First-best toll	Second-best toll	Third-best toll	Step toll	Uniform toll	No-toll equilibrium
Arnott et al., 1991	Two-point distribution	√	-	-	-	-	-	-	-	√
Arnott et al., 1996, 1999	General distribution	√ ²	-	√	-	-	-	-	-	√
Lindsey, 1994, 1999	General & two-point distribution	√ ²	-	-	√	-	-	-	-	√
Long et al., 2022	General distribution	√	-	-	-	√	√	√	-	√
Xiao et al., 2015	Uniform distribution	√	-	-	-	-	√	√	-	√
Jiang et al., 2021	General distribution	-	√	√	-	-	-	-	√	√
Jiang et al., 2022	Uniform distribution	√	-	-	-	-	-	√	-	√
Zhu et al., 2018	Uniform distribution	-	√	√	√ ¹	-	-	-	√	√
Yu et al., 2023a	Two-point distribution	√	-	√	-	-	-	-	√	√
Chu, 1999	None	-	-	√	√	-	-	√	√	√
Mun, 1999, 2003	None	-	-	√	√	-	-	-	-	√
This paper	Uniform distribution	√	-	√	√	√	√	√	√	-

Note: ¹ For a more limited “exogenous” distribution of free-flow travel time, Zhu et al. (2018) do analyse a time-variant toll that works as in the deterministic model and is first-best in that setting.

² These authors also consider uncertain demand.

Zhu et al. (2018) analysed time-varying tolling under price-sensitive demand using a bottleneck model with uncertain free-flow travel time that does not affect queuing. This leads to an outcome similar to that in the deterministic setting, as queuing can be fully eliminated. In contrast, in our model, the social optimum has queuing in “bad” states. It also misses the interaction between queuing and uncertainty. All this makes their model more tractable, but arguably less realistic, and yielding different policy implications as it makes tolling appear better for welfare and less harmful for consumers.

Yu et al. (2023a) considered information provision and uniform tolling under uncertain bottleneck capacity and price-sensitive demand. Lam (2000), Jiang et al. (2021) and Zhu et al. (2018) analysed uniform tolling in a bottleneck model with an uncertain free-flow travel time. Xiao et al. (2015), Jiang et al. (2022) and Long et al. (2022) also considered single-step tolling under fixed demand. The literature that is most directly related to our study looks at time-variant tolling in the stochastic bottleneck model with uncertain capacity but with fixed demand. This literature yields the first-, second- and third-best tolls as discussed before (see Lindsey, 1994, 1996, 1999; Xiao et al., 2015 and Long et al., 2022). Schrage (2006) studied time-variant tolling under dynamic flow congestion and uncertain capacity. Finally, Zhang et al. (2018) studied a bottleneck model where the capacity drops by a random amount if congestion gets severe enough, and they analysed time-varying and step tolling. Their model is similar to that of Zhu et al. (2018) – who used an uncertain free-flow travel time – in that optimal tolling will remove all queuing. This is not true under our uncertain capacity, thereby complicating the analysis and making tolling less beneficial.

There are various forms that uncertainty can take in the bottleneck. Earlier studies have looked at: i) uncertain capacity (e.g. Arnott et al., 1991; Xiao et al., 2015; Long et al., 2022; Jiang et al., 2022); ii) uncertain demand (e.g. Fosgerau, 2010); iii) both uncertain capacity and demand (e.g. Arnott et al., 1993b, 1996, 1999); iv) uncertain arrival times at the bottleneck (e.g. Daniel, 1995); v) uncertain free-flow travel time (e.g. Zhu et al., 2018; Lam, 2000; Siu and Lo, 2009; Jiang et al., 2021); and vi) uncertainty in the demand function such as a random demand intercept (e.g. Fu et al., 2018). For uncertainty in the capacity, most papers assume—just like we do—that the capacity is uncertain but its realised value is constant throughout the peak. An exception is Fosgerau and Lindsey (2013), who studied random incidents that temporarily block the bottleneck. They analysed the user equilibrium and social optimum under fixed demand. Hall and Savage (2019) considered endogenous shocks to the capacity, and their probability increases with the traffic flow. Peer et al. (2010) and Schrage (2006) investigated capacity changes within the peak. So, for example, an accident that is clearer after an hour or a brief rain shower.

We follow much of the literature in considering a uniform distribution for our uncertainty, as this allows for more closed-form results. Xiao et al. (2014a, 2014b, 2015), Jiang et al. (2022) and Zhang et al. (2018) also used a uniform distribution. Arnott et al. (1991), Liu et al. (2020) and Yu et al. (2020, 2023a) used two-point distributions, which seems more restrictive than a continuous distribution. Lindsey (1994, 1999), Long et al. (2022), Jiang et al. (2021) and Liu et al. (2023) relaxed the assumptions and considered more general distributions.

Most papers—just like us—assume that drivers are rational and consider their expected price. So, we abstain from considering risk aversion or bounded rationality. Li et al. (2008) considered risk aversion by also adding the standard deviation of travel time to the user cost function, thus not only considering the expected travel time and schedule delays. Siu and Lo (2009), Liu and Liu (2018) and de Palma and Fosgerau (2013) also considered risk aversion. In Liu et al. (2020) and Jiang et al. (2022),

users considered a linear combination of the mean cost and the cost variation. Zhu et al. (2019) considered bounded rationality.

Fully time-variant tolls are practically impossible to implement in reality. More realistic toll schedules are uniform tolls that are constant over the day or step tolls with one or a few discrete steps in the toll. In the deterministic bottleneck model, Arnott et al. (1993a), Laih (1994), Lindsey et al. (2012) and Ren et al. (2016) proposed four different equilibrium models to examine such schemes. They differ in how to ensure that the generalised price is the same before and after the toll is lowered at time t . The ADL model of Arnott and co-authors has a mass departure for those arriving after t . The Laih model has separate queues for arrival before and after t that do not interact. In the Braking model of Lindsey and co-authors, the first drivers that will arrive after t brake and completely block the road for a while to prevent having to pay the (higher) toll or be overtaken. Finally, Ren et al. (2016) developed a model in between the Laih and braking model, where there are separate queues but the queue for arrivals after t hinders the other drivers while not fully blocking the road. Van den Berg (2012) extended these models by adding price-sensitive demand and found that more steps can increase welfare gain and make consumers better off. Whilst considering uncertainty, Xiao et al. (2015), Long et al. (2022), Jiang et al. (2022) and Zhang et al. (2018) considered single-step tolling, but they used a fixed demand. The first three papers used the ADL equilibrium model, and the last one used the Laih model.

We will study the three proposed time-varying tolls for the bottleneck model with uncertain capacity whilst adding price-sensitive demand. As we will see, this complicates the analysis and has important effects. We will also study uniform tolling and step tolling using the ADL equilibrium model. We will use a uniform distribution of the service time of the bottleneck—i.e., the inverse of capacity—and this uncertain capacity varies over the days but is constant throughout the peak.

3. Model set-up

We assume that the bottleneck capacity is constant within a day but changes stochastically from day to day. Define the “service time” of the bottleneck as $\phi = 1/s$, where s is the capacity. The service time follows a uniform distribution over an interval $[\phi_{min}, \phi_{max}]$, and $f(\phi) = 1/(\phi_{max} - \phi_{min})$ is the probability density function of ϕ . This simplifies the mathematics but does make equations less intuitive. Commuters do not know the realisation of capacity on a given day. From their day-to-day travel, they learn the capacity distribution and make their departure time choices by minimising their expected generalised price (Arnott et al., 1993b, 1996; Lindsey, 1999; Xiao et al., 2015; Long et al., 2022). As discussed in the literature review, a uniform distribution for the uncertainty is common in the literature. Extending our setting to a more general distribution seems interesting for future work.

In the classic bottleneck model (Vickrey, 1969), commuters travel from home to work through the bottleneck. Without loss of generality, we assume that the free-flow travel time is zero. Thus, the travel time departing at time t equals the queuing time at the bottleneck, $q(t, \phi)$, where $1/\phi$ is the realised capacity. Let $\omega(t)$ denote the maximum service time so that no queue exists at departure time t . Then, the queuing time at t is:

$$q(t, \phi) = \begin{cases} \phi \int_{\hat{t}}^t r(x) dx - (t - \hat{t}), & \phi > \omega(t), \\ 0, & \phi \leq \omega(t) \end{cases} \quad (1)$$

where \hat{t} is the time at which the queue begins to increase from zero and $r(t)$ is the departure rate that is the same for all capacity realisation as people depart without knowing what the capacity will be. The

travel time cost for commuters departing at t is:

$$TT(t, \phi) = \alpha q(t, \phi). \quad (2)$$

Define the schedule delay cost for commuters departing at t as:

$$SDC(t + q(t, \phi) - t^*) = \beta \max\{t^* - (t + q(t, \phi)), 0\} + \gamma \max\{(t + q(t, \phi)) - t^*, 0\}, \quad (3)$$

where α , β and γ are the values of time, schedule delay early and late, respectively. The t^* is the desired arrival time. Since capacity is stochastic, commuters departing at the same moment each day may experience different costs on different days. The expected travel cost for departure at t is:

$$E(C(t)) = \int_{\phi_{min}}^{\omega(t)} SDC(t) f(\phi) d\phi + \int_{\omega(t)}^{\phi_{max}} [TT(t, \phi) + SDC(t + q(t, \phi))] f(\phi) d\phi, \quad (4)$$

where $f(\phi)$ is the probability density function of ϕ , and $\omega(t)$ is the maximum service time for which no queue exists at t . The generalised price includes the travel cost and the toll:

$$P_j(t) = E(C_j(t)) + \tau_j(t), \quad (5)$$

where j indicates the scenario. When $j = NT$, we have no tolling and $\tau_{NT} = 0$. We also consider a uniform toll—indicated by UT —that is constant throughout the peak and a step toll—indicated by ST —that varies in two discrete steps. Finally, we have three fully time-varying tolls: our first-best toll indicated by FB ; the second-best toll indicated by SB with a constant departure rate (as taken from Long et al. (2022)); and the third-best toll indicated by TB that keeps the generalised price the same as in the NT and has a constant departure rate (as taken from Xiao et al. (2015)).

The inverse demand function is $D(N)$, where N is the total demand. In user equilibrium, we have $P_j = D(N_j)$. The welfare or social surplus is:

$$SS_j = \int_0^{N_j} D(n) dn - TC_j, \quad (6)$$

where TC_j is scenario j 's total expected travel cost of $E(C_j(t)) \cdot N_j$.

4. Social optimum under price-sensitive demand and capacity uncertainty

In the deterministic bottleneck model, the queue can be eliminated by a time-varying toll, and this achieves the social optimum in which the generalised price is unchanged compared to the no-toll equilibrium. As we will see, the toll in our stochastic bottleneck model will affect the price and demand.

Under the time-varying scheme, the social surplus or welfare is:

$$\max_{t_s, t_e, r_{FB}(t), N_{TV}} SS_{FB} = \int_0^{N_{FB}} D(n) dn - \int_{t_s}^{t_e} E(C_{FB}(t, \phi)) r_{FB}(t) dt, \quad (7)$$

where t_s , t_e and $r_{FB}(t)$ denote the first departure time, the latest departure time and the first-best departure rate at t , respectively. Then, the generalised price follows Eq. (5) with $j = FB$. We solve for the first-best social optimum in two stages. In the first stage, for a given demand, we optimise the departure rate by using dynamic optimisation and minimising the expected total social cost, i.e. the second term of Eq. (7). This implies how the toll should change over time. This stage is similar to Lindsey (1994, 1999) in using optimal control theory,³ except for our service time following a uniform

³ Yang and Huang (1997) used optimal control theory to analyse the deterministic bottleneck model, Mun (1999) to analyse his flow congestion model, Schrage (2006) and Yu et al. (2023b) to analyse dynamic flow congestion.

distribution. The stage is also similar to Fosgerau and Lindsey (2013). In the second stage, we optimise total travel demand.

4.1. First stage: analytics of optimising the departure rate and toll development

In the dynamic optimisation, the departure rate $r_{FB}(t)$ is the control variable. $R(t)$ is the state variable denoting the cumulative departures at t : $dR(t)/dt = r_{FB}(t)$. The queuing time $q(t, \phi)$ from Eq. (1) is a second state variable. In the first stage, the problem can be reformulated as

$$\min_{t_s, t_e, r_{FB}(t)} \int_{t_s}^{t_e} E(C_{FB}(t, \phi)) r_{FB}(t) dt, \quad (8)$$

subject to

$$\frac{dq(t, \phi)}{dt} = \begin{cases} \phi r_{FB}(t) - 1, & \phi > \omega(t) \\ 0, & \phi \leq \omega(t) \end{cases} \quad (\text{Costate variable is } \mu_1(t, \phi)), \quad (9)$$

$$\frac{dR(t)}{dt} = r_{FB}(t) \quad (\text{Costate variable is } \mu_2(t)), \quad (10)$$

where $R(t_s) = 0$ and $R(t_e) = N_{FB}$. The t_s and t_e are the first and latest departure times, respectively. Note that the queue development depends on the capacity realisation, but the departure rate does not.

Let $\mu_1(t, \phi)$ and $\mu_2(t)$ denote the ‘‘costate variables’’ of $q(t, \phi)$ and $R(t)$, respectively. We set up the equations so that μ_2 is the marginal social cost (MSC) of the total departures at t : how much *higher* the total cost will be when the total departures at t are *higher*. Similarly, $\mu_1(t, \phi)$ is the shadow cost of queuing time when service time is ϕ (and thus capacity $1/\phi$): it gives how much higher the total costs are when there is more queuing.

The Hamiltonian function for the optimisation problem is:⁴

$$\begin{aligned} H(t) = & -r_{FB}(t) \left\{ \int_{\omega(t)}^{\phi_{max}} [\alpha q(t, \phi) + SDC(t + q(t, \phi))] f(\phi) d\phi + \int_{\phi_{min}}^{\omega(t)} SDC(t) f(\phi) d\phi \right\} + \\ & \int_{\omega(t)}^{\phi_{max}} \mu_1(t, \phi) (\phi r_{FB}(t) - 1) f(\phi) d\phi + \mu_2(t) r_{FB}(t). \end{aligned} \quad (11)$$

The equations of motion of costate variables $\mu_1(t, \phi)$ and μ_2 are:⁵

$$\frac{d\mu_1(t, \phi)}{dt} = -\frac{\partial H}{\partial q} = \begin{cases} -r_{FB}(t) \left(\alpha + \frac{dSDC(t+q(t, \phi))}{dq(t, \phi)} \right), & \text{if } \phi \geq \omega(t) \\ 0, & \text{if } \phi \leq \omega(t) \end{cases} \quad (12)$$

$$\frac{d\mu_2(t)}{dt} = -\frac{\partial H}{\partial R} = 0. \quad (13)$$

The transversality conditions at the start and end of the peak are $H(t_s) = 0$ and $H(t_e) = 0$.

Costate variable $\mu_1(t, \phi) \geq 0$ is the shadow cost of queuing time. It not only depends on departure time t but also on the capacity realisation and thus ϕ . At any t , if ϕ is smaller (i.e. the realised capacity is larger), queuing is shorter or even absent. At t_e , all commuters have left, and no later departures⁶ would be delayed by queuing regardless of the realised capacity. This implies that no matter what the capacity is, the shadow cost of queuing at t_e is zero:

⁴ The Hamiltonian function may look different than expected as we need to integrate over the uncertainty and have two outcomes: one with queuing and one without. Although Eq. (10) shows that there is a regime shift between when there is queuing and when there is not, there is no regime shift in the control variable of the departure nor in the toll, as these depend on the expected outcome before the capacity is known. This is also why the first constraint for queuing is integrated over ϕ , while the second for the (total) departures is not. Therefore $\mu_1(t, \phi)$ depends on the capacity realisation and thus ϕ , whereas μ_2 does not.

⁵ Where the $q(t, \phi)$ in $\frac{d\mu_1(t, \phi)}{dt} = -\frac{\partial H}{\partial q(t, \phi)}$ it is the effect for a specific realisation of ϕ , as the queuing development is specific for each ϕ .

⁶ Of course, one would delay the departures at t_e , but this effect is of size zero since one integrates from t_e to the same time t_e .

$$\mu_1(t_e, \phi) = 0. \quad (14)$$

From (12), $\mu_1(t, \phi)$ weakly decreases over time, since more commuters have departed and fewer will be delayed by queuing. Therefore, $\mu_1(t, \phi) \geq 0$, $t \in [t_s, t_e]$.

From (13), the costate variable μ_2 is constant over time. This implies that the marginal social cost is constant between t_s and t_e and equals μ_2 . The first-order condition for the departure rate r_{FB} implies:

$$\begin{aligned} \mu_2 = & \int_{\omega(t)}^{\phi_{max}} [\alpha q(t, \phi) + SDC(t + q(t, \phi))] f(\phi) d\phi + \int_{\phi_{min}}^{\omega(t)} SDC(t) f(\phi) d\phi + \\ & \int_{\omega(t)}^{\phi_{max}} \phi \mu_1(t, \phi) f(\phi) d\phi, \quad t \in [t_s, t_e]. \end{aligned} \quad (15)$$

Here, the first two terms in integrals together give the expected private travel cost for a departure at t , and thus the expected marginal external cost (MEC) at t equals the last term $\int_{\omega(t)}^{\phi_{max}} \phi \mu_1(t, \phi) f(\phi) d\phi$, which is the expected cost due to queuing imposed on later departures at t . Moreover, the transversality conditions and (15) imply that the MEC is zero at t_s and t_e , and non-negative in between.

Now, let us search for the toll's pattern. For the expected price to be constant over time and thus the users to be in user equilibrium, the toll must vary just like the expected MEC:

$$\tau(t) = \int_{\omega(t)}^{\phi_{max}} \phi \mu_1(t, \phi) f(\phi) d\phi + \tau_0, \quad (16)$$

where τ_0 is the starting level of the toll at t_s . And the toll must again equal τ_0 at t_e .

We can also use this to derive the optimal departure rate and how it varies over time. The optimal departure rate must equal:

$$r_{FB}(t) = 1/\omega(t), \quad t \in [t_s, t_e], \quad (17)$$

where again $1/\omega(t)$ is the minimum of the stochastic capacity for which there is no queue at t . The intuition is that once the queue starts to develop, it does not collapse to zero until the last departure, as stated in Lindsey (1994, 1999). Suppose $t' \in [t_s, t_e]$ and $r(t') > 1/\omega(t')$; there would always be an interval that the expected delay cannot dissipate efficiently. Conversely, if $r(t') < 1/\omega(t')$, there would always be an interval in which capacity is not fully used. Moreover, for our uniform distribution, we get the following departure rates at the start and end of the peak:

$$\begin{aligned} r_{FB}(t_s) &= 1/\phi_{max}, \\ r_{FB}(t_e) &= (\alpha + \gamma)/(\alpha\phi_{max} + \gamma\phi_{min}). \end{aligned} \quad (18)$$

So, the departure rate at the start of the peak is low and equals the minimum capacity of $1/\phi_{max}$. The departure rate ends high, being above the ‘‘average’’ capacity of $1/\bar{\phi} = 2/(\phi_{max} + \phi_{min})$ but below the maximum capacity. As $\omega(t)$ is a continuous, but possibly kinked, function with a uniform distribution, it also follows that the optimal departure rate will continuously increase over departure times during at least a part of the peak and never decreases over departure times. There may be departure windows when drivers always or never see queuing, which implies that the departure rate $r_{FB}(t)$ are constant over time in those ranges.

The departure rate starts low, as it is more costly if an early departure causes queuing if the realised capacity turns out to be low. The optimal departure rate ends high, as for late departures there are few to no users who can be affected. This is also implied by the pattern of shadow cost of queuing: $\mu_1(t, \phi)$.

4.2. Second stage: setting total demand

Our first-stage results with a uniform distribution are consistent with those of Lindsey (1994, 1999)

for a general distribution, but our explicit distribution allows for more explicit results. We now turn to the second stage where we set the total number of users to maximise the reduced-form social surplus (which is conditional on the departure rate following the socially optimal pattern):

$$SS_{FB} = \int_0^{N_{FB}} D(n) dn - N_{FB} E(C_{FB}^*(N_{FB})),$$

where the superscript * in C_{FB}^* indicates that it is the travel cost under the socially optimal departure rate from the previous section. Maximising this to N_{FB} gives

$$D(N_{FB}) = E(C_{FB}^*(N_{FB})) + N_{FB} \frac{\partial E(C_{FB}^*(N_{FB}))}{\partial N_{FB}} = \mu_2. \quad (19)$$

And the inverse demand should equal the (expected) marginal social cost that equals the μ_2 from the previous section. This in turn implies that the toll should start at zero, and thus the first-best toll is:

$$\tau(t) = \int_{\omega(t)}^{\phi_{max}} \phi \mu_1(t, \phi) f(\phi) d\phi. \quad (20)$$

The toll at t should equal the expected $MEC(t)$.

This outcome is, of course, similar to the social optimum with a fixed capacity, where it is also optimal that the inverse demand equals the MSC and the toll equals $MEC(t)$. However, in that case, the optimal departure rate equals the fixed capacity, whereas here the rate starts at the minimum capacity and then weakly increases to some final level that is in between the ‘average’ and the maximum.

In the optimum, the peak is divided into different time windows: $[t_s, t_1]$, $[t_1, t_2]$, $[t_2, t^*]$ and $[t^*, t_e]$, denoted as “*Situations 1–4*”, respectively.⁷ Commuters have a different queuing experience and schedule delay experience under the given capacity (day) in different time windows. The type of queuing and schedule delay commuters may experience, and the generalised price during each time window, are given in Appendix A.1.

From the analysis above, we have the following results:

Remark 1: The cumulative departures $R(t_1) = \frac{t_1 - t_s}{\phi_{max}}$, $R(t_2) = \frac{t^* - t_s}{\phi_{max}}$ and $R(t^*) = \frac{t_1 - t_s}{\phi_{max}} + \frac{t^* - t_1}{\omega(t^*)}$.

The proof of Remark 1 is given in Appendix A.2.

Remark 2: In the social optimum, we have:

$$t_s = \frac{2(\alpha+\gamma)(\beta+\gamma)\phi_{max}t^* - \gamma^2(\phi_{max} - \phi_{min})t_1}{2(\alpha+\gamma)(\beta+\gamma)\phi_{max} - \gamma^2(\phi_{max} - \phi_{min})} - \frac{\gamma\phi_{max}[\gamma(\phi_{max} + \phi_{min}) + 2\alpha\phi_{max}]N_{FB}}{2(\alpha+\gamma)(\beta+\gamma)\phi_{max} - \gamma^2(\phi_{max} - \phi_{min})} \quad (21)$$

$$t_e = \frac{2(\beta+\gamma)(\alpha\phi_{max} + \gamma\phi_{min})t^* + (2\beta+\gamma)\gamma(\phi_{max} - \phi_{min})t_1}{2(\alpha+\gamma)(\beta+\gamma)\phi_{max} - \gamma^2(\phi_{max} - \phi_{min})} + \frac{2\beta\phi_{max}(\alpha\phi_{max} + \gamma\phi_{min})N_{FB}}{2(\alpha+\gamma)(\beta+\gamma)\phi_{max} - \gamma^2(\phi_{max} - \phi_{min})} \quad (22)$$

The proof of Remark 2 is given in Appendix A.3.

The tolls at the first and at the latest departure time are both zero. The generalised price of the first commuter is $P_{FB} = \beta(t^* - t_s)$. This is also the marginal social cost, since an additional user departing before t_s does not impose any costs on others.

⁷ In the no-toll equilibrium, it could occur that the last departure is before t^* if the uncertain capacity is really spread out (Arnott et al., 1996). But in the social optimum there will always be departures after t^* . The departure rate is positive between t_s and t_e , non-decreasing over t and starts at the minimum capacity. If the last departure occurred before t^* someone departing at t_e could depart marginally later; they would lower their cost without hurting anyone else and thus lower the total cost. If capacity turns out high, they will face no travel delay whilst lowering their schedule delay; if capacity turns out low, they will join the queue for the last departures and keep the same schedule delay whilst lowering queueing time (Lindsey, 1994). Hence, all four situations will occur.

To summarise, compared with the deterministic model, both the generalised price and the total demand will change under the toll. Both departure rate and toll are now non-linear over time. The departure rate starts low at the start of the peak as it equals the minimum capacity; at the end of the peak, the rate is high but below the maximum capacity; in between these times, the departure rate is non-decreasing, continuous and increasing for a range of departure times. Interestingly, this is the mirror image of the pattern without tolling: the untolled departure rate starts high, decreases in between and ends low. All this is really different from the deterministic bottleneck model, where first-best tolling leads to a constant departure rate and has the same price as the no-toll case.

4.3. Analytical comparison with the no-toll equilibrium and other proposed time-varying tolls

For the no-toll equilibrium, we use the results from Arnott et al. (1996, 1999), Lindsey (1994), Xiao et al. (2015) and Long et al. (2022), as our focus is on tolled equilibria. The analysis may actually be more complicated without tolling, as there are multiple cases of the no-toll equilibrium depending on the parameters, e.g. with or without departure after t^* , whereas in the social optimum there are always departures after t^* no matter what the parameters.

By adopting the results of Long et al. (2022), the first and the last departure time under the no-toll equilibrium when ϕ follows the uniform distribution are given as follows:

(a) when there is departure after t^* ,

$$t_s^{NT} = t^* - \frac{\alpha+\gamma}{\beta+\gamma} N \left[\frac{\phi_{min}+\phi_{max}}{2} - \frac{\alpha(\alpha\phi_{max}+\gamma\phi_{min})+(\alpha+\gamma)\alpha\phi_{min}}{(\alpha+\gamma)^2(\phi_{min}+\phi_{max})} \right] \quad (23)$$

$$t_e^{NT} = t^* + N \left[\frac{\alpha\phi_{max}+\gamma\phi_{min}}{(\alpha+\gamma)} - \frac{(\alpha+\gamma)(\phi_{min}+\phi_{max})}{2(\beta+\gamma)} + \frac{\alpha(\alpha\phi_{max}+\gamma\phi_{min})+(\alpha+\gamma)\alpha\phi_{min}}{(\alpha+\gamma)(\beta+\gamma)(\phi_{min}+\phi_{max})} \right]; \quad (24)$$

(b) when there is no departure after t^* ,

$$t_s^{NT} = t^* - N\hat{\phi} \quad (25)$$

$$t_e^{NT} = t^*, \quad (26)$$

where $\hat{\phi}$ is obtained by solving the equation $\frac{1}{\hat{\phi}} \left(\frac{\phi_{min}+\phi_{max}}{2} - \frac{(\hat{\phi})^2 - (\phi_{min})^2}{(\phi_{max})^2 - (\phi_{min})^2} \right) + \frac{\hat{\phi} - \phi_{min}}{\phi_{max} - \phi_{min}} = \frac{\alpha+\beta+\gamma}{\alpha+\gamma}$.

By comparing Eqs. (21)–(22) and Eqs. (23)–(26), it becomes apparent that the departure timings under the social optimum differ from those under the no-toll equilibrium. Then, the price also changes under the social optimum, which is different from the results in the deterministic scenario.

The departure rate in the no-toll stochastic equilibrium is qualitatively the mirror image of the socially optimal one. The untolled departure rate starts high, weakly decreases in between and ends low. Numerical analysis in Xiao et al. (2015) shows that the rate may be flat in some periods, which we will also see in our numerical model for the first-best toll. We cannot analytically compare the prices in the no-toll and first-best equilibria. We will do this in the numerical model and find that no matter the parameters, the first best has higher prices than the no toll.

All this is also very different from the deterministic bottleneck model, where the no-toll rate is flat with a downward jump for arrivals at t^* . Moreover, the no-toll and first-best settings have the same prices in the deterministic bottleneck model. The first-best toll in the stochastic model is convex over time, where it starts and ends at zero. Conversely, in the deterministic model, the toll is piecewise linear. Our stochastic bottleneck model is thus more akin to flow congestion models (e.g. Agnew, 1973; Chu, 1999; Mun, 2003).

Now, let us turn to the comparison with the two other time-variant toll models in the literature. Long et al. (2022) proposed a time-varying toll under the condition that the departure rate is constant

over time. This has the advantages of being easier to implement and allowing for a more detailed analytical study. However, as the socially optimal rate varies (see also Linsley, 1994, 1999), it must mean that their scheme has a (somewhat) lower welfare. Hence, we call it the “second-best” time-variant toll, as the constraint of a constant departure rate somewhat lowers welfare. Xiao et al. (2015) proposed a time-variant toll that has a constant departure rate and keeps the price at the no-toll setting. It thus adds another constraint to the tolling; hence, we call it “third best”. As it is a constrained version of the second best, it can, at most, do as well as the second best (when the second best would lead to the same price as no tolling). Again, analytical comparison is limited, but in the numerical model, we see that the third-best toll always has lower prices than the second best and first best but has a lower social surplus. It is thus nicer for the users, making it more politically feasible to implement. All three time-variant tolls are non-linear and concave over time.

5. Uniform toll

The uniform toll is constant during the morning peak, and it does not give an incentive to change the departure patterns other than an indirectly via changing total demand. Thus, the equilibrium is the same as without tolling (Arnott, 1996; Long et al., 2022). However, these authors did not study flat tolling. Yu et al. (2023a) did look at uniform tolling.

The generalised price follows from Eq. (5) with $j = UT$. The social surplus is:

$$SS_{UT} = \int_0^{N_{UT}} D(n) dn - N_{UT}E(C_{UT}). \quad (27)$$

Maximising it implies that the flat toll needs to equal the marginal external cost (MEC):

$$\tau_{UT} = N_{UT} \frac{\partial E(C_{UT})}{\partial N_{UT}}. \quad (28)$$

The derivations are given in Appendix A.4. The optimal demand N_{UT} is found by $P_{UT} = D(N_{UT})$. This is basically the same outcome as the flat toll without uncertainty (Arnott et al., 1993a). This tolling regime has many possible cases depending on the parameters, just like the case without tolling.

This uniform toll cannot alter departure rates, it can only lower the total number of users so that the (averaged over time) MSC equals the inverse demand, whereas without tolling, the inverse demand equals the travel cost, which is lower than the MEC. Hence, the uniform toll raises the expected price and lowers the total number of users.

6. Single-step toll

The single-step toll consists of a “time-variant” step-up component that applies during the step-tolling period, and a time-invariant component that lasts the whole peak:

$$\tau_{ST}(t) = \rho_t + \mu. \quad (29)$$

Here, $\tau_{ST}(t)$ is the toll at departure time t , and ρ_t is the time-variant step part implemented from t^+ to t^- . The t^+ and t^- are the start time and end time of the step-tolling period. μ is the time-invariant part implemented throughout the peak.

Adapting the procedure in Van den Berg (2012), we again optimise in two stages. In the first stage, given the number of users, the time-variant part ρ_t and the step-tolling period are obtained by minimising the total expected cost. In the second stage, the total demand is set to maximise social welfare whilst considering the effect on the first stage. This then also implies the time-invariant part of

the toll: μ .

With fixed demand, Long et al. (2022) investigated the single-step toll in the stochastic bottleneck model, where the service time of the bottleneck follows a general distribution. If the step toll is implemented, commuters depart at a constant departure rate before t^+ . When the toll is lifted at t^- , there will be a mass of commuters departing. Like Long et al. (2022), we use the ADL (Arnott et al., 1993a) model with mass departures.

In the first stage, the results are the same as in Long et al. (2022) for a uniformly distributed service time. There are two cases under the single-step toll, depending on whether the peak ends at or after the preferred arrival time t^* . Here, under the single-step toll, the peak ends after t^* in Case I (it is Case 8.2 in Long et al. (2022)), and it ends at t^* in Case II (it is Case 9.2 in Long et al. (2022)). For the second stage, Appendix A.5 gives the detailed derivations, and the results are summarised below.

Following Long et al. (2022), the expected total social cost in Case I is:

$$TC_{ST}(N_{ST}) = -\frac{W(\vec{\phi})(M(\vec{\phi}))^2(N_{ST})^2}{4(\beta+\gamma)^2} + \frac{(\alpha+\gamma)(\bar{\phi}-\tilde{\phi})\beta(N_{ST})^2}{\beta+\gamma}. \quad (30)$$

Here, $\vec{\phi}$ is the reciprocals of the average departure rate during the period from the departure time of the first commuter to the departure time of the first commuter who pays the toll, and $\vec{\phi}$ is obtained by solving the equation $M(\vec{\phi})\frac{dW(\vec{\phi})}{d\vec{\phi}} + 2W(\vec{\phi})\frac{dM(\vec{\phi})}{d\vec{\phi}} = 0$, where $W(\vec{\phi}) = \frac{1}{Y(\vec{\phi}) + \frac{2}{(\alpha+\gamma)\vec{\phi}}}$, $Y(\vec{\phi}) = \frac{1}{\alpha\vec{\phi} - (\alpha-\beta)H(\vec{\phi})}$,

$$H(\vec{\phi}) = \bar{\phi} - G(\vec{\phi}) + \vec{\phi}F(\vec{\phi}), \quad M(\vec{\phi}) = \beta(\alpha+\gamma)(\bar{\phi}-\tilde{\phi})Y(\vec{\phi}) - \beta(\beta+\gamma)H(\vec{\phi})Y(\vec{\phi}) + (\beta+\gamma) + \frac{2\beta(\bar{\phi}-\tilde{\phi})}{\vec{\phi}}, \quad F(x) = \frac{x-\phi_{min}}{\phi_{max}-\phi_{min}}, \quad G(x) = \frac{x^2-\phi_{min}^2}{2(\phi_{max}-\phi_{min})}, \quad \tilde{\phi} = G\left(\frac{\alpha\phi_{max}+\gamma\phi_{min}}{\alpha+\gamma}\right) \text{ and } \bar{\phi} = \frac{\phi_{max}+\phi_{min}}{2}.$$

These definitions follow from those in Long et al. (2022). As proved in Proposition 16 by Long et al. (2022), $\vec{\phi}$ is independent of N_{ST} . Then, the total expected social cost shown in Eq. (30) is a function of travel demand. The total toll revenue is:

$$TR_{ST} = \rho_t N_1 + \mu N_{ST}, \quad (31)$$

where $\rho_t = \frac{M(\vec{\phi})W(\vec{\phi})N_{ST}}{2(\beta+\gamma)}$ is the optimal time-variant step part of the toll, and $N_1 = N_{ST} - \frac{\rho_t}{W(\vec{\phi})}$ is the number of travellers departing during the step-tolling period. The step part of the toll, ρ_t , is a function of travel demand. The average marginal external cost (MEC) is the difference between the average marginal social cost and the average travel cost. From Eq. (30), the average MEC is:

$$MEC_{ST} = -\frac{W(\vec{\phi})(M(\vec{\phi}))^2 N_{ST}}{4(\beta+\gamma)^2} + \frac{(\alpha+\gamma)(\bar{\phi}-\tilde{\phi})\beta N_{ST}}{\beta+\gamma}. \quad (32)$$

We obtain the optimal time-invariant part of the toll by equalising the average marginal external cost and the average toll, i.e. $MEC_{ST} = \frac{TR_{ST}}{N_{ST}}$. Then, the optimal time-invariant part is:

$$\mu = MEC_{ST} - \frac{\rho_t N_1}{N_{ST}} = \frac{(\alpha+\gamma)(\bar{\phi}-\tilde{\phi})\beta N_{ST}}{\beta+\gamma} - \frac{W(\vec{\phi})M(\vec{\phi})N_{ST}}{2(\beta+\gamma)}. \quad (33)$$

The generalised price in Case I is:

$$P_{ST} = -\frac{W(\vec{\phi})(M(\vec{\phi}))^2 N_{ST}}{2(\beta+\gamma)^2} + \frac{2(\alpha+\gamma)(\bar{\phi}-\tilde{\phi})\beta N_{ST}}{\beta+\gamma}. \quad (34)$$

From Eq. (34), the optimal demand under a step toll can be found by $P_{ST} = D(N_{ST})$. The results

for Case II will be similarly obtained and are given in Appendix A.6 to save space.

7. Numerical study

Because the analytical results are not easy to interpret, we conduct a numerical study to compare the different schemes. This also provides new insights beyond what can be obtained from the analytics. We conduct extensive sensitivity tests to see how robust these insights are. We will use a uniformly distributed bottleneck service time as this is a common assumption, as seen in the literature review.

The first section will look at fixed demand, and the second at price-sensitive demand. Fixed demand allows for a clearer comparison of departure rates and costs, as otherwise the number of users will differ between tolling regimes, which muddles the view.

Unless otherwise mentioned, we use the same values for the unit cost parameters as Long et al. (2022): $\alpha = 6.4\$/h$, and ratios $\beta/\alpha = 0.609$ and $\gamma/\beta = 3.9$. The desired arrival time $t^* = 9h$. Following Long et al. (2022), the mean bottleneck service time is $\bar{\phi} = 1s/veh$, which implies an ‘‘average’’ capacity of $3600\ veh/h$. Increasing ϕ_{min} (the lower bound of service time) makes the system less uncertain. When ϕ_{min} approaches the average $\bar{\phi}$, the model approaches the deterministic scenario. In this set-up, $\phi_{max} = 2 - \phi_{min}$.

The inverse demand function is assumed to be linear: $D(n) = d_0 - d_1n$. The average demand elasticity is -0.4 (Van den Berg, 2012). When the capacity is at the mean value (i.e. the capacity is $1/\bar{\phi}$), the number of commuters $N = 5000\ veh$. under the no-toll scheme (Long et al., 2022). This implies that $d_0 = 15.0893$ and $d_1 = 0.0022$. The demand elasticity is set based on the mean (average) capacity and $N = 5000\ veh$ under this setting. Due to the uncertainty, the demand and price will change under the no-toll equilibrium when the elasticity changes.

7.1. Numerical evaluation of the different scenarios under fixed demand

Long et al. (2022) proposed a ‘‘second-best’’ time-varying toll model with two properties: (i) the toll keeps the departure rate constant; (ii) there are no toll charges for both the first and the last traveller, as is the case in the deterministic bottleneck model. Xiao et al. (2015) introduced a ‘‘third-best’’ time-varying toll that has these conditions and keeps the generalised price (i.e. the sum of travel cost and toll) at the level in the no-toll equilibrium. In these previous works, the demand is fixed, and the time-varying toll scheme cannot achieve the system optimum. The ‘‘first-best’’ toll of Lindsey (1999) does this and leads to a departure rate that weakly increases over time.

For different ϕ_{min} , Fig. 1 plots the equilibrium departure rates under the no-toll and the first-, second- and third-best time-varying toll schemes. Note that the departure rate in no-toll equilibrium follows from Long et al. (2022).

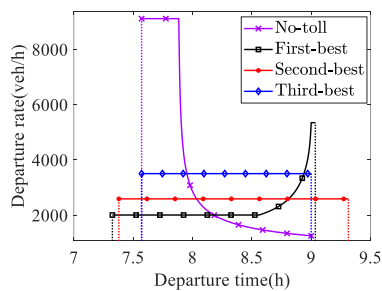
The departure rate in no-toll equilibrium is non-increasing over the departure time, while that under the first-best toll is non-decreasing. The intuition for the latter is that as time progresses, more travellers have departed and fewer travellers will be delayed by queuing, which means that the cost of queuing decreases over time in the optimum. Thus, in the earliest departure period, the departure rate is constant and equals the minimum capacity to avoid high queuing cost; after that, it starts to increase, before becoming constant again for the last departure period. During the early peak, capacity is underutilised unless the worst possible traffic condition occurs. The results are consistent with those in Lindsey (1994), who used a binary distribution, and in Fosgerau and Lindsey (2013), who used a capacity that varies over the day. In line with the analytics, Fig. 1 shows that the optimal departure rate is constant after t^* ,

and is in between the ‘average’ capacity and the maximum capacity. It is interesting to point out that when ϕ_{min} is relatively large with less uncertainty, the first-best departure rate after t^* is the same as that under the third-best toll, as shown in Fig. 1(c)–(d).

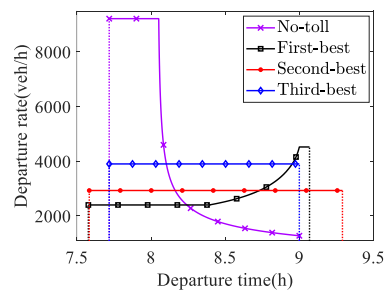
The third-best time-varying toll keeps the price unchanged from the no-toll case, and the peak duration and timing are also the same. With our uniform distribution, the peak begins earlier and ends later in the social optimum compared to that under the no-toll equilibrium. Under the second-best time-varying toll, the peak is the longest due to a lower departure rate on average. Conversely, the third-best time-varying toll has the shortest departure window. Thus, compared with the no-toll case, both the first- and second-best tolls increase the price and the peak duration under fixed demand, which differs from the deterministic model.

The no-toll departure rate is qualitatively the mirror image of the first-best one. It weakly decreases over time and is continuous with two kinks. The rate starts at the maximum, and then, after a while, it starts to decrease.

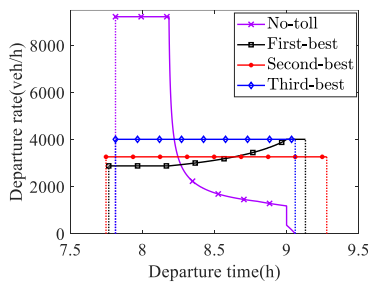
When ϕ_{min} increases towards $\bar{\phi}$, so that there is less uncertainty, the departure rate under the first-best toll becomes more concentrated and that under the three time-varying toll schemes approaches the mean capacity, as illustrated in Fig. 1(d). A constant departure rate is only optimal when there is no uncertainty.



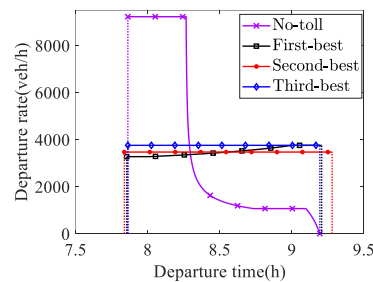
(a) High uncertainty: $\phi_{min}=0.2$ s/veh.



(b) Medium uncertainty: $\phi_{min}=0.5$ s/veh.



(c) Low uncertainty: $\phi_{min}=0.75$ s/veh.



(d) Very low uncertainty: $\phi_{min}=0.9$ s/veh.

Fig. 1: Comparisons of equilibrium departure rates under different time-varying tolling schemes in stochastic bottleneck model for four spreads of the capacity.

Note: The uniform distribution of the service time varies from ϕ_{min} to $\phi_{max} = 2 - \phi_{min}$. So, lowering ϕ_{min} increases the spread of the distribution without altering the mean.

We also examine the results for a different ratio β/α , where the ratio γ/β is kept constant. For simplicity, the results are illustrated in Fig. B.1 in Appendix B. We find that when the ratio of β/α decreases, there is a decrease in the overall departure rates, and the difference under different schemes

becomes small. Since the cost of queuing becomes relatively high, the average departure rate decreases to avoid being involved in long queues due to uncertainty. Also, the peak under these schemes begins earlier with a smaller ratio of β/α , since experiencing schedule delays becomes less costly.

For the three toll schemes, Fig. 2 shows the mean queuing cost and mean schedule delay cost over departure time. All tolls start at zero. The third-best toll is much lower than the others and its peak is much shorter, which results in much higher expected travel times. In the graph, the first-best toll is higher than the second best. But this result is not universal and depends on the parameter levels: with low uncertainty, we obtain the opposite result. Nevertheless, the first-best toll does start the earliest, regardless of the parameters. For simplicity, the results when $\phi_{min}=0.8$ s/veh are given in Fig. B.2.

Queuing will not be eliminated under uncertainty. It does not occur during the earliest departure period in the social optimum, which is consistent with the analytics. However, after that, queuing is always possible. For the second- and third-best toll schemes, because of the constant departure rate, the mean queuing cost increases linearly over departure time. In the social optimum, expected travel time delays are low for much of the peak, but they increase sharply during the later peak because of the sudden increase in the departure rate then. This increase is more pronounced the more uncertain the capacity is. The mean schedule delay cost at the end of the peak is smaller than that at the beginning, as tolling will not eliminate all delays (since then, the departure rate has never exceeded the minimum capacity, and almost always some capacity will go unused).

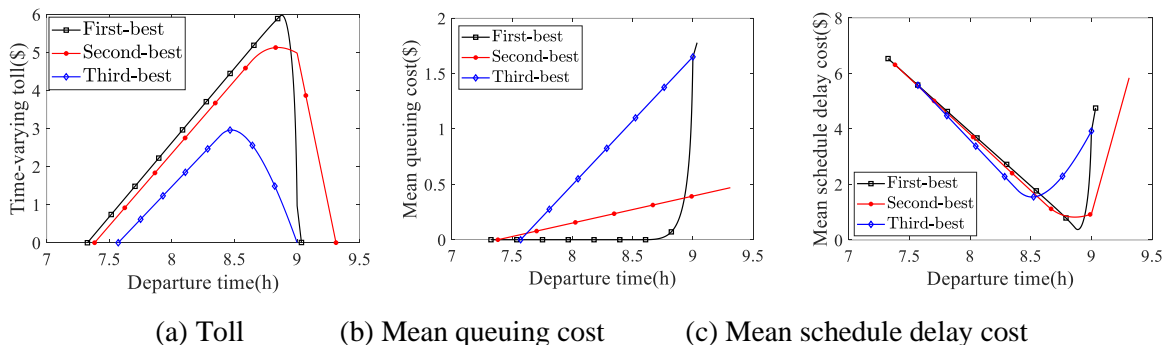


Fig. 2: Comparisons of toll, mean queuing cost and mean schedule delay when $\phi_{min}=0.2$ s/veh.

Fig. 3 illustrates the effect of tolling over different degrees of uncertainty. When ϕ_{min} increases, there is less uncertainty, and the price, average travel cost, average queuing cost and schedule delay cost decrease. Unlike in the deterministic model, the first-best toll raises the generalised price from the untolled setting, and it decreases travel cost less in percentage terms. These effects are stronger the more uncertainty there is.

As mentioned above, the third-best time-varying toll keeps the generalised price unchanged from the no-toll equilibrium. In Fig. 3(a), both the first- and second-best tolls raise the price. When there is more uncertainty, the first-best price is higher than the second-best price. The expected average travel cost under the optimal toll is naturally at the minimum, while this is not necessarily true for queuing costs. The expected queuing cost under the second-best scheme is lower than that under the first-best one, as illustrated in Fig. 3(c). Fig. 3(d) shows that the average schedule delay cost under the social optimum is the lowest and that under the second-best time-varying toll is the highest.

Together with the analysis above, we give some possible intuitive explanations concerning the

results under the first- and second-best tolls. As shown in Fig. 1, the departure duration is longest and the constant departure rate is lowest under the second-best toll. The second-best toll spreads the departures greatly, so that the expected queuing cost is lowest while the expected schedule delay cost is highest. This is also shown in Fig. 3(c) and (d). Compared to the first-best toll, the second-best toll is more effective at eliminating queues and less so at reducing schedule delay costs. Moreover, both uncertainty and the ratio β/α could affect the two schemes. Comparatively, the second-best toll is more sensitive to the ratio β/α , while the first-best toll is more sensitive to uncertainty.

Specifically, with a relatively small ratio β/α , experiencing schedule delay is less costly. The second-best toll always begins earlier, resulting in a higher price than the first-best toll, no matter whether ϕ_{min} is relatively small or large (also see Fig. B.1 in Appendix B). Conversely, with a relatively large ratio β/α , experiencing schedule delay is more costly, and the second-best toll begins either earlier or later than the first-best toll depending on the uncertainty (see Fig. 1). This results in either a higher or a lower generalised price of the second-best toll (see Fig. 3(a)) compared to the first-best one. When ϕ_{min} is relatively small, the first-best toll begins earlier, leading to a higher price. Under that circumstance, the toll is higher and the constant component of the departure rate (in the earliest part of the peak) lasts longer under the first-best scheme in a highly uncertain environment (see Fig. 2(a) and 1(a)–(b)). The intuition is that travellers could avoid the possibility of a long queue caused by large uncertainty and being heavily penalised by a high cost of arriving late. When there is less uncertainty, the price under the first-best toll decreases faster. When ϕ_{min} increases and exceeds a certain threshold, the price under the first-best toll is lower than the second-best price. Similar results are observed in the next section when there is price-sensitive demand.

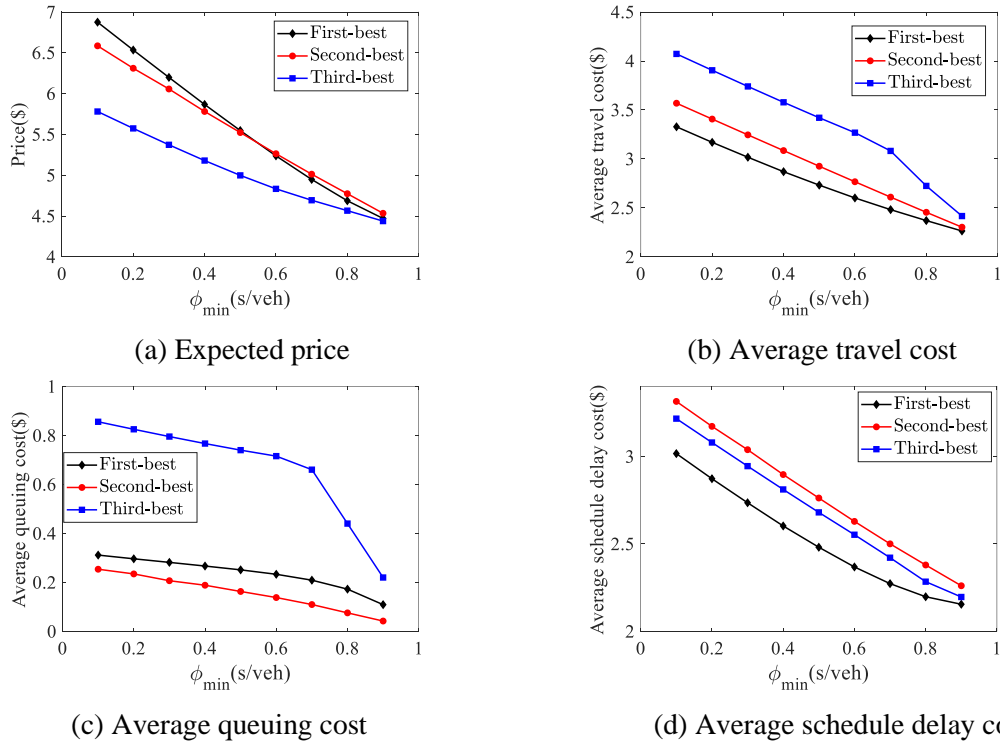


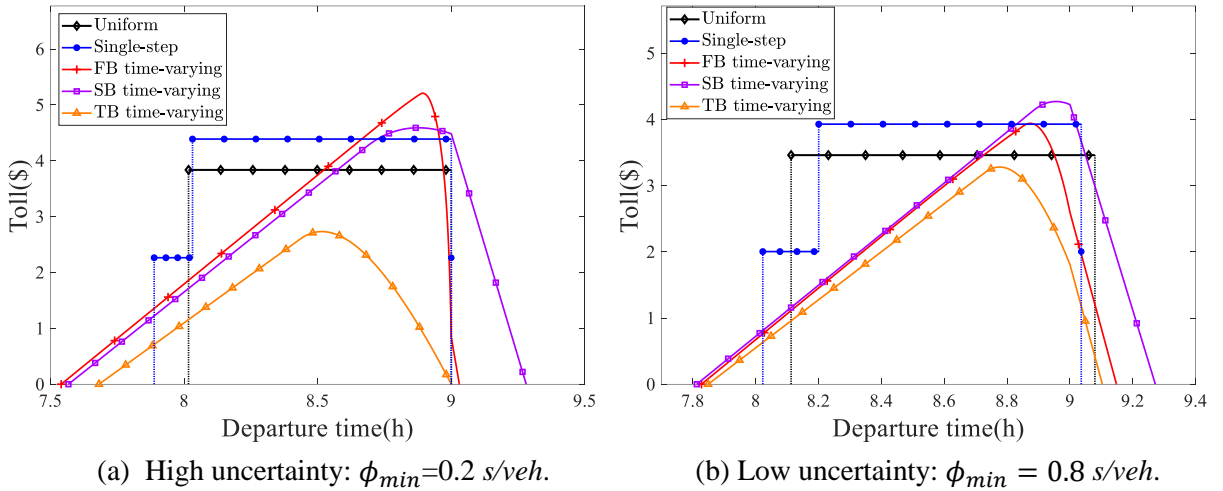
Fig. 3: The effect of uncertainty under different time-varying toll schemes with fixed demand on (a) price, (b) average travel cost, (c) average queuing cost, and (d) average schedule delay cost.

7.2. Numerical evaluation of the different scenarios under price-sensitive demand

This section compares different tolling schemes. Here, the second- and third-best time-varying tolls mentioned in Section 7.1 are generalised to the price-sensitive schemes involving optimisation of demand as a separate step. Specifically, with price-sensitive demand, the second-best time-varying toll scheme imposes a constant departure rate. The third-best time-varying scheme adds the second constraint that the price is the same as in the no-toll equilibrium. We find their optimal departure rates by numerically maximising the social welfare with respect to the departure rate.

7.2.1. The structures of different tolling schemes

For two levels of uncertainty, Fig. 4 plots the uniform toll, single-step toll, first-best (FB) time-varying toll, second-best (SB) time-varying toll and third-best (TB) time-varying toll with price-sensitive demand. The flat toll and single-step toll are generally higher than the average toll of the time-varying schemes. Similarly to the results for the deterministic model, the marginal external cost is higher with coarser pricing, implying higher toll levels. Furthermore, for the single-step toll, the time-variant step part of the toll is lifted at some point and there is a mass departure of users and they only need to pay the time-invariant part of the toll. Finally, similarly to what we found for fixed demand, with more uncertainty, the first-best time-varying toll begins earlier and the average toll is higher than for the second-best time-varying toll; the results are opposite with less uncertainty.



(a) High uncertainty: $\phi_{min}=0.2$ s/veh.

(b) Low uncertainty: $\phi_{min} = 0.8$ s/veh.

Fig. 4: Different tolling schemes over the departure time for two levels of uncertainty.

7.2.2. The effect of uncertainty

In order to compare the efficiency of different tolling schemes with price-sensitive demand, we employ the index ω_i^{SS} that denotes the relative efficiency of tolling scheme i , i.e. the social surplus gain of scheme i from the no-toll equilibrium relative to that of the first-best scheme. It equals

$$\omega_i^{SS} = \frac{SS_i - SS_{NT}}{SS_{FB} - SS_{NT}},$$

where SS_{NT} , SS_{FB} and SS_i denote the social surplus of the system under the no-toll equilibrium, under the first-best equilibrium and under the tolling scheme i , respectively.

Fig. 5 shows how uncertainty affects the demand, price, expected average travel cost, average toll, social surplus and relative efficiency under different tolling schemes. There are kinks in the curves for the third-best toll when the equilibrium pattern changes.

We make the following observations. First, travel demand increases and the generalised price decreases with respect to ϕ_{min} . The demand is highest and the price is lowest under the third-best time-varying toll, where they are the same as under the no-toll equilibrium. With more uncertainty, the first- and second-best time-varying schemes raise the price more from the no-toll equilibrium, thus reducing the demand much lower. There is only a moderate difference between the first- and second-best time-varying schemes. Similarly to the results with fixed demand, the price is higher and demand is lower under the first-best toll than under the second-best one when ϕ_{min} is less than around 0.5. The results are the opposite when ϕ_{min} is relatively large. Compared with time-varying schemes, a flat toll and step toll raise the price and reduce demand more.

Second, from Fig. 5(c)–(d), the average travel cost and toll decrease with ϕ_{min} . When there is less uncertainty, the average travel cost and toll under the third- and second-best schemes approach those under the first-best toll. Conversely, those under the single-step toll and flat toll become closer to that of the first-best toll as the degree of uncertainty increases. Additionally, the average travel cost under the third-best toll is higher than that of the single-step toll when ϕ_{min} is less than a critical value, of around 0.7, and it decreases sharply when ϕ_{min} exceeds the critical value. Fig. 5(d) shows that the average toll under the third-best scheme is always lowest.

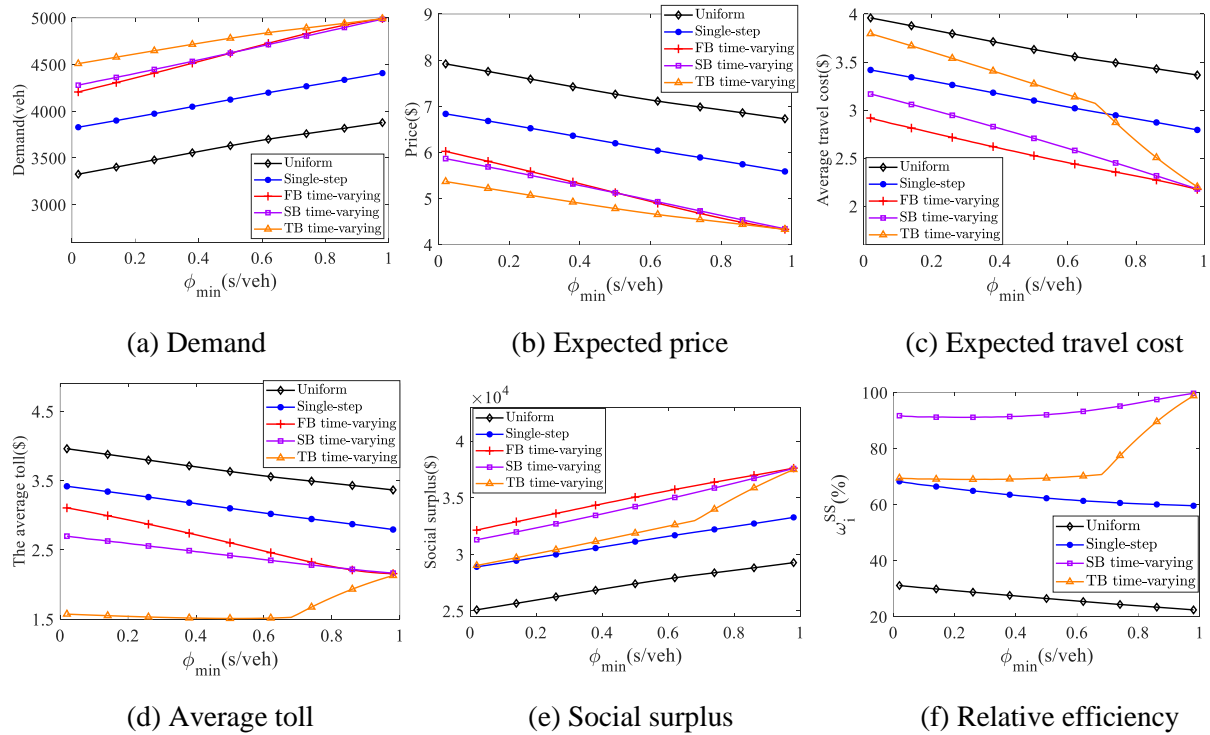


Fig. 5: The effect of uncertainty on the outcomes under price-sensitive demand.

Note: The relative efficiency of policy i is $\omega_i^{SS} = \frac{SS_i - SS_{NT}}{SS_{FB} - SS_{NT}}$, where SS_i is the social surplus or welfare.

Finally, Fig. 5(e) shows that the social surplus of the third-best time-varying scheme is always higher than that of the single-step scheme, even though the average cost might be higher under the third-best toll. According to Fig. 5(f), with the increases of ϕ_{min} the second- and third-best time-varying schemes become more efficient. The three time-varying schemes approach the same one when the stochastic bottleneck model approaches the deterministic model. However, the relative efficiency of the flat toll and single-step toll is higher with uncertainty than without uncertainty. This indicates that

lowering demand is more important in a highly uncertain environment, since the step toll is better at changing demand levels than lowering delays. The relative efficiency of the second-best toll is the highest, followed by the third-best toll, the single-step toll and then the uniform toll.

7.2.3. The effect of demand elasticity

Fig. 6 illustrates how demand elasticity affects the demand, average travel cost and relative efficiency of our schemes when $\phi_{min} = 0.2$. The results with less uncertainty ($\phi_{min} = 0.75$) are given in Fig. B.3 in Appendix B. When the elasticity changes from -0.2 to -2, demand becomes more sensitive to the price. Compared to time-varying schemes, the performance of step-toll schemes is more sensitive to elasticity. The demand under the third-best toll is by definition the same as in the no-toll case.

According to Fig. 6, more price-sensitive demand means a lower demand and average travel cost. Furthermore, there is little difference in demand between the first- and second-best schemes. Demand under the third-best toll scheme is always largest. This might also result in the highest travel cost, especially when demand is more price-sensitive to the price.

Finally, Fig. 6(c) illustrates that as demand becomes more price-sensitive, the relative efficiency of the single-step toll and uniform toll increases, while that of time-varying schemes remains almost constant. The efficiency of the single-step toll is higher than that of the third-best time-varying toll when demand elasticity is less than around -0.8. By contrast, when there is less uncertainty, the uniform and single-step tolls become less efficient, and their performance is inferior to time-varying schemes (see Fig. B.3 in Appendix B). This is because these schemes are better at changing the total demand and worse at changing departure times. When demand becomes more price-sensitive, the step toll becomes more efficient since it is more important to change demand then. However, with less uncertainty in the system, the time-varying tolls perform better because the shift in the travel window matters more.

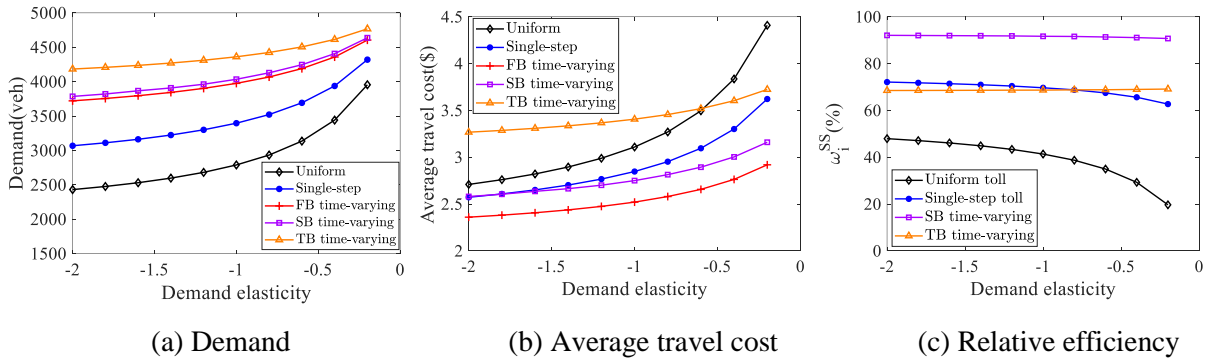


Fig. 6: The effect of elasticity on the outcomes when $\phi_{min} = 0.2$ s/veh.

Note: The relative efficiency of policy i gives the relative welfare gain compared to the first best and is $\omega_i^{SS} = \frac{SS_i - SS_{NT}}{SS_{FB} - SS_{NT}}$, where SS_i is the social surplus or welfare.

7.2.4. The effect of β/α

How the ratio β/α affects the price, average travel cost and relative efficiency under those tolling schemes is illustrated in Fig. 7. The results with less uncertainty ($\phi_{min} = 0.8$) are given in Fig. B.4 in Appendix B. Here, the value of time $\alpha = 6.4$ \$/h and the ratio of γ/β remain unchanged. Note that ratio β/α varies from 0.26 to 0.99 to ensure we do not violate $\beta < \alpha$. With the increase of β/α , queuing becomes less costly, while experiencing schedule delay becomes more costly.

With the increase of β/α , the price and average travel cost increase due to the increase in value of schedule delay, as illustrated in Fig. 7(a). Regardless of β/α , the performance of the second-best toll is

close to that of the first best. The price under the third-best time-varying toll is again always lowest.

Furthermore, with a lower value of schedule delay, the average travel cost might be higher under the third-best scheme than that under the uniform toll (see Fig. 7(b)). Under these circumstances, queuing is relatively costly, indicating that the third-best scheme is worse at reducing queue delays.

Finally, in Fig. 7(c), with the increase of β/α , the third-best time-varying toll becomes more efficient, which also indicates that the impact of uncertainty on its performance becomes less important. However, the efficiency of the single-step toll slightly increases, and that of the other two schemes remains almost constant. When the system becomes more certain, the efficiency under all schemes becomes almost insensitive to β/α (see Fig. B.4 in Appendix B).

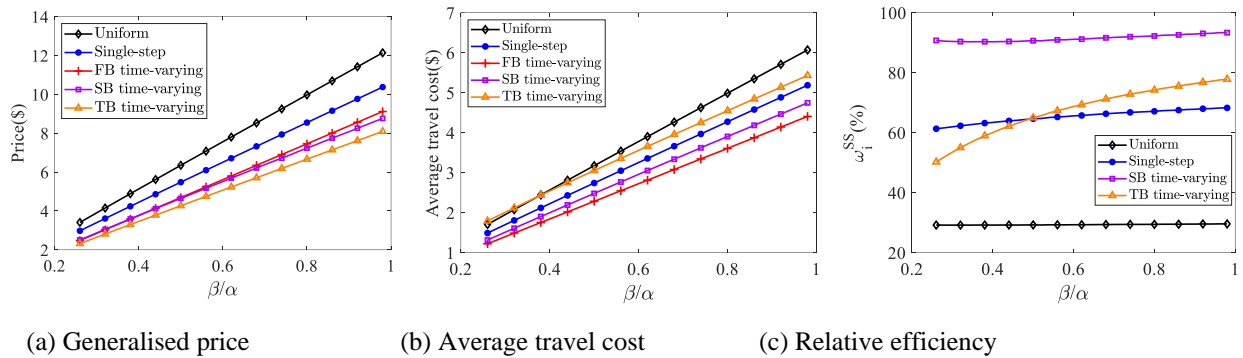


Fig. 7: The effect of β/α on the outcomes when $\phi_{min} = 0.2$ s/veh.

Note: The relative efficiency of policy i is $\omega_i^{SS} = \frac{SS_i - SS_{NT}}{SS_{FB} - SS_{NT}}$, where SS_i is the social surplus or welfare.

7.3. Summarising the numerical analysis

This section compares the socially optimal time-varying toll to a second-best and third-best toll that, by assumption, have a departure rate that is constant over time. The third-best toll adds a further constraint that the generalised price should be the same as without tolling. We also looked at a uniform toll—which is constant over time—and a single-step toll.

In the social optimum, the departure rate weakly increases over the morning; in the early peak, it is constant over time and equals the minimum capacity. Thereafter, it increases over time, and then finally, in the latest part of the peak, it is constant again. This differs greatly from the deterministic model where the optimal departure rate is always constant over time. Conversely, the untolled outcome has a rate that is the mirror image and weakly decreases over time. Unlike in the deterministic model, the first-best toll raises the generalised price from the untolled setting, and it decreases travel cost and increases welfare less in percentage terms. These effects are stronger the more uncertainty there is.

The second-best toll attains a welfare level that is somewhat lower than in the first-best one, whereas the third-best toll attains a much lower welfare. The relative efficiencies of the second-best and third-best schemes fall with the degree of uncertainty, whereas the price sensitivity has little to no effect on relative efficiency. The uniform toll has a welfare that is much lower than that of the single-step toll. For most parameter ranges, the step toll, in turn, has a lower welfare than the third-best toll; exceptions include when demand is very price-sensitive, and when the value of time is substantially higher than that of schedule delay early.

As demand becomes more price-sensitive, all schemes result in generalised prices that are closer together, and under perfectly elastic demand, they would all lead to the same price. The uniform toll has the highest price, followed by the single-step toll, then either the first-best or second-best toll, and finally

the third-best toll, which has the same generalised price as the no-toll case. Whether the first- or second-best toll has the higher price depends on multiple parameters, but their prices will be close. The uniform and step toll attain welfare levels closer to the first-best level when the price sensitivity increases.

8. Conclusion

This study examines various tolling schemes in the stochastic bottleneck model with price-sensitive demand. We assume that the capacity is constant within a day but changes stochastically from day to day. We study three time-varying toll schemes, a uniform toll and a single-step toll. Our core contribution is doing so under uncertain capacity and price-sensitive demand. Previous works only looked at these separately, and their interplay changes the results substantially. This is important for policymaking, since their combined occurrence is what might be expected in reality: travel time uncertainty is a fact of life in real transport systems, while travellers may respond to this and to policies such as tolling by changing their departure time, but also by changing their behaviour in other ways (including mode choice, working from home or rescheduling more drastically to times outside the peak).

In the stochastic bottleneck model, the first-best social optimum is decentralised by a time-varying toll and has a departure rate that is continuous and weakly increases over time: the rate is constant in the earliest and latest parts of the peak, while in between the rate increases. This is very different from the deterministic model, where the optimal rate is constant over time. We solve for the time-varying tolls and the step toll in two stages. In the first stage, under a given demand, the departure rate is optimised, and this in turn implies the toll development over time. In the second stage, the total demand is optimised, which implies the toll's starting level. We also study a second-best and a third-best time-varying toll, which are based on schemes in earlier papers but now account for price-sensitive demand. The second-best scheme assumes that the departure rate should be constant over time, as is optimal in the deterministic bottleneck model but is second best here; this imposition leads to a lower welfare than under first-best tolling. The third-best time-varying scheme adds to the further condition that the generalised price should be the same as in the no-toll case. This is optimal in the deterministic model but lowers welfare under uncertainty.

Our numerical study illustrates all of this. It shows that the optimal departure rate weakly increases over the morning, whereas the untolled departure rate is the mirror image as the rate weakly decreases. The second-best and third-best tolls are optimal only when there is no uncertainty. Unlike in the deterministic model, the first-best toll raises the generalised price from the untolled setting, and it decreases travel cost and raises welfare less in percentage terms. These effects are stronger the more uncertainty there is. The second-best toll attains a welfare a bit below that of the first-best toll, whereas the third-best toll leads to a much lower welfare and a higher travel cost. The single-step toll has a welfare that is below that of the third-best toll for most parameter ranges; exceptions include when demand is very price-sensitive and when the value of time is much higher than that of schedule delay early. The uniform toll has a welfare that is well below that of the step toll. When there is less uncertainty, the second-best and third-best schemes lead to a welfare that is closer to that of the first-best one, whereas the uniform and single-step toll perform relatively worse with less uncertainty. The effects of the price elasticity on the uniform toll and single-step toll are more significant than on the time-varying tolls. All this shows the importance of jointly considering uncertainty and price-sensitive demand.

Our study can be extended in several directions. The first one is to consider other step-toll models, as we used the ADL step toll. How would adding more steps to the step toll alter things? All the time-

varying tolls vary non-linearly over time, whereas the deterministic model has a linear slope. How would such a much simpler and easier-to-implement linear toll perform?⁸ What toll would a profit-maximising road operator set? In the deterministic model, the operator uses the same toll pattern as the welfare-maximising first-best toll but with a time-invariant markup added. Does the same hold with stochasticity? What happens if we consider larger networks or a different form of congestion? What if demand is also uncertain or the uncertainty varies over the day (e.g. an accident that is cleared)? What about alternative policies to tolling, such as capacity expansion or travel credits? Finally, accounting for the effect of information is important (Yu et al., 2021, 2023a; Han et al., 2021; Verhoef et al., 1996). So, there is a long list of interesting and important opportunities for further research on stochastic and congested transportation systems.

Finally, as a policy conclusion, we find that in the stochastic bottleneck model, unlike the deterministic one, the first-best socially optimal toll cannot remove all queuing and hurts consumers by raising the generalised price from the untolled case. Moreover, the percentage welfare gain is much lower in our stochastic model than in the deterministic bottleneck model. This makes tolling harder to implement politically. Adding a constraint that the generalised price must remain the same as without tolling is beneficial for users (before considering revenue recycling); however, this raises travel costs and lowers toll revenue and welfare. This highlights the fact that the interaction between uncertainty and price sensitivity complicates the design of transportation policies and alters their effects; and, of course, in reality we do have uncertainty, making it important to consider it.

Acknowledgement

Financial support from China Scholarship Council (202007090099) is gratefully acknowledged. RJ acknowledges the support by the National Natural Science Foundation of China (Grant No. 72288101, 71931002).

References

- Agnew, C.E., 1977. The theory of congestion tolls. *Journal of Regional Science*, 17(3), 381-393.
- Arnott, R., de Palma, A., Lindsey, R. 1991. Does providing information to drivers reduce traffic congestion?. *Transportation Research Part A*, 25(5), 309-318.
- Arnott, R., de Palma, A., Lindsey, R., 1993a. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *Am. Econ. Rev.*, 83 (1), 161-179.
- Arnott, R., A. de Palma, Lindsey, 1993b. Information and stochastic capacity and demand in the bottleneck model. Working paper (November).
- Arnott, R., de Palma, A., Lindsey, R., 1996. Information and usage of free-access congestible facilities with stochastic capacity and demand. *Inter. Econ. Rev.*, 37, 181-203.
- Arnott, R., de Palma, A., Lindsey, R., 1999. Information and time-of-usage decisions in the bottleneck model with stochastic capacity and demand. *Eur. Econ. Rev.*, 43, 525-548.
- Daniel, J.I., 1995. Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues. *Econometrica* 63(2), 327-370.
- de Palma, A. and Fosgerau, M., 2013. Random queues and risk averse users. *European Journal of Operational Research*, 230(2), 313-320.
- Chu, X., 1999. Alternative congestion technologies. *Regional Science and Urban Economics* 29 (6), 697-722.

⁸ Chu (1999) analyses this for his dynamic flow congestion model without uncertainty.

- Fosgerau, M., 2010. On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. *J. Econ. Dyn. Control* 34(4), 598–603
- Fosgerau, M., Lindsey R., 2013. Trip-timing decisions with traffic incidents. *Regional Science and Urban Economics*, 43, 764-782.
- Fu, X., van den Berg, V.A.C., Verhoef, E. T., 2018. Private road networks with uncertain demand. *Research in Transportation Economics*, 70, 57-68.
- Jiang, G.G., Lo, H.K., Tang, Q.R., Liang Z., Wang, S.L., 2021. The impact of road pricing on travel time variability under elastic demand, *Transportmetrica B: Transport Dynamics*, 9(1), 595-621.
- Jiang, G.G., Wang, S.L., Lo, H.K., Liang, Z., 2022. Modeling cost variability in a bottleneck model with degradable capacity, *Transportmetrica B: Transport Dynamics*, 10(1), 84-110.
- Hall, J.D., Savage, I., 2019. Tolling roads to improve reliability. *Journal of Urban Economics* 113, 103187.
- Han, X., Yu, Y., Gao, Z.Y., Zhang, H.M., 2021. The value of pre-trip information on departure time and route choice in the morning commute under stochastic traffic conditions. *Transportation Research Part B*, 152, 205-226.
- Klein, I., Levy, N., Ben-Elia, E., 2018. An agent-based model of the emergence of cooperation and a fair and stable system optimum using ATIS on a simple road network. *Transportation research part C: emerging technologies*, 86, 183-201.
- Laih, C.H., 1994. Queuing at a bottleneck with single and multi-step tolls. *Transportation Research Part A*, 28 (3), 197-208.
- Lam, T.C., 2000. The effect of variability of travel time on route and time-of-day choice. Ph D dissertation, Department of Economics, University of California at Irvine.
- Li, H., Bovy, P.H. and Bliemer, M.C., 2008. Departure time distribution in the stochastic bottleneck model. *International Journal of ITS Research*, 6(2) 79–86.
- Li, Z.C., Huang, H.J., Yang, H., 2020. Fifty years of the bottleneck model: A bibliometric review and future research directions. *Transportation Research Part B* 139, 311-342.
- Li, M., Jiang, G., Lo, H.K., 2022. Pricing strategy of ride-sourcing services under travel time variability. *Transportation Research Part E: Logistics and Transportation Review*, 159, 102631.
- Liang, Z., Jiang, G., Lo, H. K., 2023. Dynamic equilibrium analyses in a ride-sourcing market under travel time uncertainty. *Transportation Research Part C: Emerging Technologies*, 153, 104222.
- Lindsey, R., 1994. Optimal departure scheduling in the morning rush hour when capacity is uncertain. Presented at the 41st North Amer. Meeting of the Regional Sci. Assoc., Niagara Falls, Ontario.
- Lindsey, R., 1996. Optimal departure scheduling for the morning rush hour when capacity is uncertain. Volume 2: Modelling Transport Systems. Proceedings of the 7th World Conference on Transport Research World Conference on Transport Research Society.
- Lindsey, R., 1999. Optimal departure scheduling for the morning rush hour when capacity is uncertain. In: Emmerink, R., Nijkamp, P. (Eds.), *Behavioural and Network Impacts of Driver Information Systems* (Vol. 12). Astigate: Aldershot, England.
- Lindsey, R., 2009. Cost recovery from congestion tolls with random capacity and demand. *J. Urban. Econ.* 66, 16-24.
- Lindsey, R., Daniel, T., Gisches, E., Rapoport, A., 2014. Pre-trip information and route-choice decisions with stochastic travel conditions: Theory. *Transportation Research Part B: Methodological*, 67, 187-207.
- Lindsey, R., van den Berg, V.A.C., Verhoef, E.T., 2012. Step tolling with bottleneck queuing congestion.

- J. Urban Econ., 72 (1), 46-59.
- Liu, P., Liu, Y., 2018. Optimal information provision at bottleneck equilibrium with risk-averse travelers. *Transportation Research Record*, 2672(48), 69-78.
- Liu, Q., Jiang, R., Liu, R., Zhao, H., Gao, Z., 2020. Travel cost budget based user equilibrium in a bottleneck model with stochastic capacity. *Transportation Research Part B: Methodological*, 139, 1-37.
- Liu, Q., Jiang, R., Liu, W., Gao, Z., 2023. Departure time choices in the morning commute with a mixed distribution of capacity. *Transportation Research Part C: Emerging Technologies*, 147, 104011.
- Long, J., Yang, H., Szeto, W.Y., 2022. Departure time equilibrium and tolling strategies of a bottleneck with stochastic capacity, *Transp. Sci.*, 56(1), 79-102.
- Long, J., Tan, W., Szeto, W.Y., Li, Y., 2018. Ride-sharing with travel time uncertainty. *Transportation Research Part B: Methodological*, 118, 143-171.
- Ma, W., Zeng, L., An, K., 2023. Dynamic vehicle routing problem for flexible buses considering stochastic requests. *Transportation Research Part C: Emerging Technologies*, 148, 104030.
- Maher, M., Stewart, K., Rosa, A., 2005. Stochastic social optimum traffic assignment. *Transportation Research Part B*, 39(8), 753-767.
- Meng, Q., Liu, Z., 2011. Trial-and-error method for congestion pricing scheme under side-constrained probit-based stochastic user equilibrium conditions. *Transportation*, 38(5), 819-843.
- Mun, S.I., 1999. Peak-load pricing of a bottleneck with traffic jam. *Journal of Urban Economics*, 46(3), 323-349.
- Mun, S.L., 2003. Bottleneck Congestion with Traffic Jam: A reformulation and correction of earlier result. Kyoto University working paper (version of February), accessed from <https://www.econ.kyoto-u.ac.jp/~mun/papers/Bottleneck0920.pdf> on 17 November 2023.
- Peer, S., Koster, P.R., Verhoef, E.T., Rouwendal, J. 2010. Traffic incidents and the bottleneck model. Working paper (version of 13 January 2014). Accessed from www.researchgate.net/profile/Stefanie-Peer/publication/229054198_Traffic_incidents_and_the_bottleneck_model/links/0c96052d3a1d55358b000000/Traffic-incidents-and-the-bottleneck-model.pdf on 21 November 2023.
- Ren, H., Xue, Y., Long, J., Gao, Z. 2016. A single-step-toll equilibrium for the bottleneck model with dropped capacity. *Transportmetrica B: Transport Dynamics*, 4(2), 92-110.
- Siu, B., Lo, H.K., 2009. Equilibrium trip scheduling in congested traffic under uncertainty. Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), *Transportation and Traffic Theory 2009: Golden Jubilee*. Hong Kong Springer, US, pp. 19-38.
- Small, K.A., 2015. The bottleneck model: An assessment and interpretation. *Economics of Transportation* 4(1-2), 110-117.
- Small, K.A., Verhoef, E.T., 2007. *The Economics of Urban Transportation*. Routledge, London.
- Schrage, A., 2006. Traffic Congestion and Accidents. *Regensburger Diskussionsbeiträge zur Wirtschaftswissenschaft*, Working Paper No. 419 (version of 9 November 2006). Accessed from https://epub.uni-regensburg.de/4535/1/Congestion_and_Accidents_WP.pdf on 21 November 2023.
- Van den Berg, V.A.C., 2012. Step-tolling with price-sensitive demand: why more steps in the toll make the consumer better off. *Transportation Research Part A* 46(10), 1608-1622.
- Verhoef, E.T., Emmerink, R.H.M., Nijkamp, P., Rietveld, P., 1996. Information provision, flat- and fine congestion tolling and the efficiency of road usage. *Reg. Sci. Urban Econ.*, 26(5): 505-530.
- Vickrey, W.S., 1969. Congestion theory and transport investment. *Am. Econ. Rev* (Papers and

- Proceedings), 59 (2), 251-261.
- Xiao, L.L., Liu, R., Huang, H.J., 2014a. Stochastic bottleneck capacity, merging traffic and morning commute. *Transportation Research Part E: Logistics and Transportation Review*, 64, 48-70.
- Xiao, L.L., Liu, R., Huang, H.J., 2014b. Congestion behavior under uncertainty on morning commute with preferred arrival time interval. *Discrete Dynamics in Nature and Society*, 2014, 767851.
- Xiao, L.L., Huang, H.J., Liu, R.H., 2015. Congestion behavior and tolls in a bottleneck model with stochastic capacity. *Transp. Sci.*, 49(1), 46-65.
- Yang, H., Huang, H.J., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. *Transportation Research Part B*, 31(6), 425-440.
- Yang, H., 1999. System optimum, stochastic user equilibrium and optimal link tolls. *Transp. Science*, 33, 354-360.
- Yu, X., van den Berg, V.A.C., Li, Z.C., 2023a. Congestion pricing and information provision under uncertainty: Responsive versus habitual pricing. *Transportation Research Part E: Logistics and Transportation Review*, 175, 103119.
- Yu, X., van den Berg, V.A.C., Verhoef, E.T., 2023b. Congestion pricing under dynamic flow congestion and heterogeneous preferences. VU working paper.
- Yu, Y., Han, X., Jiang, R., Darr, J., Jia, B., 2020. Departure time and route choices with accurate information under binary stochastic bottleneck capacity in the morning commute. *IEEE Access*, 8, 225551-225565.
- Yu, Y., Han, X., Jia, B., Jiang, R., Gao, Z.Y., Zhang, H.M., 2021. Is providing inaccurate pre-trip information better than providing no information in the morning commute under stochastic bottleneck capacity? *Transp. Res. C*, 126, 103085.
- Zhang, W.W., Zhao, H., Jiang, R., 2018. Impact of capacity drop on commuting systems under uncertainty. *Journal of Advanced Transportation* 2018, Article ID 6809304.
- Zhang, F., Lu, J., Hu, X., Fan, R., Chen, J., 2022. Managing bottleneck congestion with tradable credit scheme under demand uncertainty. *Research in Transportation Economics*, 95, 101232.
- Zhu, S., Jiang, G.G., Lo, H.K., 2018. Capturing value of reliability through road pricing in congested traffic under uncertainty. *Transportation Research Part C*, 94, 236-249.
- Zhu, Z., Li, X., Liu, W., Yang, H., 2019. Day-to-day evolution of departure time choice in stochastic capacity bottleneck models with bounded rationality and various information perceptions. *Transportation Research Part E*, 131, 168-192.

Appendix A. Analytics for different toll schemes

A.1 The generalised price in each time window under social optimum

Situation 1: $[t_s, t_1]$. During the period, commuters always experience schedule delay early and never experience queuing regardless of the realised capacity. The departure rate $r_{FB}(t) = 1/\phi_{max}$, where $t \in [t_s, t_1]$. The generalised price during $[t_s, t_1]$ is:

$$P_{FB}(t) = \beta(t^* - t) + \tau_{FB}(t) \quad (\text{A.1})$$

Situation 2: $[t_1, t_2]$. During the second period, commuters always experience schedule delay early. They experience queuing for some realisations of capacity. They experience queuing if $q(t, \phi) > 0$, i.e., $\phi > \omega(t)$; otherwise, they do not. The generalised price during $[t_1, t_2]$ is:

$$P_{FB}(t) = \beta(t^* - t) + (\alpha - \beta) \int_{\omega(t)}^{\phi_{max}} q(t, \phi) f(\phi) d\phi + \tau_{FB}(t) \quad (\text{A.2})$$

Situation 3: $[t_2, t^*]$. Commuters possibly experience schedule delay either early or late, and possibly experience queuing depending on the realised capacity. They experience schedule delay late with queuing if $t + q(t, \phi) > t^*$, i.e., $\phi > \frac{t^* - t_1}{R(t) - (t_1 - t_s)/\phi_{max}}$; they experience schedule delay early with queuing if $t + q(t, \phi) < t^*$, i.e., $\omega(t) < \phi < \frac{t^* - t_1}{R(t) - (t_1 - t_s)/\phi_{max}}$; they experience schedule delay early without queuing if $\phi < \omega(t)$. The generalised price during $[t_2, t^*]$ is:

$$P_{FB}(t) = \int_{\phi_{min}}^{\omega(t)} \beta(t^* - t) f(\phi) d\phi + \int_{\omega(t)}^{\frac{t^* - t_1}{R(t) - (t_1 - t_s)/\phi_{max}}} [\beta(t^* - (t + q(t, \phi))) + \alpha q(t, \phi)] f(\phi) d\phi + \int_{\frac{t^* - t_1}{R(t) - (t_1 - t_s)/\phi_{max}}}^{\phi_{max}} [\gamma(t + q(t, \phi) - t^*) + \alpha q(t, \phi)] f(\phi) d\phi + \tau_{FB}(t) \quad (\text{A.3})$$

Situation 4: $[t^*, t_e]$. Commuters always experience schedule delay late regardless of the capacity, and experience queuing for some realisations of capacity. They experience queuing if $q(t, \phi) > 0$, i.e., $\phi > \omega(t)$, otherwise, they do not. The generalised price during $[t^*, t_e]$ is:

$$P_{FB}(t) = \int_{\phi_{min}}^{\omega(t)} \gamma(t - t^*) f(\phi) d\phi + \int_{\omega(t)}^{\phi_{max}} [\gamma(t + q(t, \phi) - t^*) + \alpha q(t, \phi)] f(\phi) d\phi + \tau_{FB}(t) \quad (\text{A.4})$$

A.2 The proof of Remark 1

Proof. The departure rate $r_{FB}(t) = 1/\phi_{max}$, where $t \in [t_s, t_1]$. Then, the cumulative departures at t_1 is $R(t_1) = (t_1 - t_s)/\phi_{max}$. From the schedule delay experience during *Situation 2* and *Situation 3*, the boundary condition for t_2 is that commuters departing at t_2 arrive exactly at t^* when the realised capacity is the minimum, i.e., $1/\phi = 1/\phi_{max}$. Then, we have $R(t_2) = (t^* - t_s)/\phi_{max}$. The cumulative departures at t^* is obtained by substituting $t = t^*$ into Eq. (A.3) and Eq. (A.4), and equalising the two equations.

A.3 The proof of Remark 2

Proof. By the definition of $\omega(t)$, we have $q(t_e, \omega(t_e)) = 0$. i.e.,

$$\omega(t_e) \left[N_{FB} - \frac{(t_1 - t_s)}{\phi_{max}} \right] - (t_e - t_1) = 0$$

In the optimum, $P_{FB}(t_s) = P_{FB}(t_e)$, then we have

$$\frac{(\alpha+\gamma)}{2} f(\phi) \left(N_{FB} - \frac{t_1-t_s}{\phi_{max}} \right) \left((\phi_{max})^2 - \omega^2(t_e) \right) - \gamma(t^* - t_1) = \beta(t^* - t_s)$$

Substituting $f(\phi) = 1/(\phi_{max} - \phi_{min})$ and $\omega(t_e) = (\alpha\phi_{max} + \gamma\phi_{min})/(\alpha + \gamma)$ into the two equations and rearrange them, the results in Remark 2 can be obtained.

A.4 Uniform toll scheme

The social surplus equals the total benefit minus total expected social cost. The total expected cost equals the expected travel cost multiplied by the number of users. Under the uniform toll, the social welfare needs to be maximised subject to the constraint that price equals the sum of mean travel cost and the toll. Therefore, the problem can be formulated as:

$$\max SS_{UT} = \int_0^{N_{UT}} D(n) dn - N_{UT} E(C_{UT}(N_{UT})) \quad (\text{A.5})$$

$$\text{s.t. } D(N_{UT}) = E(C_{UT}(N_{UT})) + \tau_{UT} \quad (\text{A.6})$$

where the subscript ‘‘UT’’ denotes the uniform toll scheme. N_{UT} , $E(C_{UT}(N_{UT}))$ and τ_{UT} denotes the demand, expected travel cost of user and toll under the uniform toll, respectively. To find the optimal constant toll τ_{UT} , the following Lagrangian is maximised,

$$L(N_{UT}, \tau_{UT}, \lambda) = \int_0^{N_{UT}} D(n) dn - N_{UT} E(C_{UT}(N_{UT})) + \lambda(D(N_{UT}) - E(C_{UT}) - \tau_{UT}) \quad (\text{A.7})$$

where λ is the multiplier. The first-order conditions are

$$\frac{\partial L}{\partial N_{UT}} = D(N_{UT}) - N_{UT} \frac{\partial E(C_{UT})}{\partial N_{UT}} - E(C_{UT}) + \lambda \left(\frac{\partial D(N_{UT})}{\partial N_{UT}} - \frac{\partial E(C_{UT})}{\partial N_{UT}} \right) = 0 \quad (\text{A.8})$$

$$\frac{\partial L}{\partial \tau_{UT}} = -\lambda = 0 \quad (\text{A.9})$$

$$\frac{\partial L}{\partial \lambda} = D(N_{UT}) - E(C_{UT}) - \tau_{UT} = 0 \quad (\text{A.10})$$

These conditions imply that the toll can be given as follows,

$$\tau_{UT} = D(N_{UT}) - E(C_{UT}) = N_{UT} \frac{\partial E(C_{UT})}{\partial N_{UT}} \quad (\text{A.11})$$

Therefore, the uniform toll is set such that the average marginal social cost equals the price, and thus the toll equals the average marginal external cost.

A.5 Single-step toll scheme

In the first stage, the step part of the toll is obtained by minimising the total social cost under the given demand. The solution is basically the same as in Long et al. (2022) with fixed demand. In the second stage, the social welfare is maximised to find the optimal flat part of toll implementing during the whole peak. The problem can be formulated as follows,

$$\max SS_{ST} = \int_0^{N_{ST}} D(n) dn - TC_{ST}(N_{ST}) \quad (\text{A.12})$$

$$\text{s.t. } D(N_{ST}) = AC_{ST}(N_{ST}) + \rho_{ST} \quad (\text{A.13})$$

where the subscript ‘‘ST’’ denotes the single-step toll scheme. TC_{ST} and AC_{ST} denote the total expected social cost and the average expected travel cost under the step part of toll, respectively, and $AC_{ST} = \frac{TC_{ST}}{N_{ST}}$.

ρ_{ST} are defined as $\rho_{ST} = \frac{\rho_t N_1}{N_{ST}} + \mu$, where ρ_t , μ , and N_1 are the time-variant step part of the toll, time-invariant part of toll and the number of users departing from t^+ to t^- , respectively.

Here, similar to the procedure in Van den Berg (2012), in the first stage, given the number of users, the time-variant part ρ_t and the step tolling period are obtained by minimising the total expected social cost. The results of TC_{ST} , ρ_t and N_1 follow those in Long et al. (2022), and TC_{ST} and ρ_t are both the function of travel demand, as given in their study. In the second stage, we also use the Lagrangian to find the solution by maximising the social welfare, since total expected travel cost and the step part of toll can be expressed as the function of travel demand. To find the optimal flat part of the toll μ , the following Lagrangian is maximised,

$$L(N_{ST}, \mu, \lambda) = \int_0^{N_{ST}} D(n) dn - TC_{ST}(N_{ST}) + \lambda(D(N_{ST}) - AC_{ST}(N_{ST}) - \rho_{ST}) \quad (\text{A.14})$$

The first-order conditions are

$$\frac{\partial L}{\partial N_{ST}} = D(N_{ST}) - \frac{\partial TC_{ST}}{\partial N_{ST}} + \lambda \left(\frac{\partial D(N_{ST})}{\partial N_{ST}} - \frac{\partial AC_{ST}}{\partial N_{ST}} - \frac{\partial \rho_{ST}}{\partial N_{ST}} \right) = 0 \quad (\text{A.15})$$

$$\frac{\partial L}{\partial \mu} = -\lambda = 0 \quad (\text{A.16})$$

$$\frac{\partial L}{\partial \lambda} = D(N_{ST}) - AC_{ST} - \rho_{ST} = 0 \quad (\text{A.17})$$

These conditions imply that the optimal flat part of the toll can be given as follows,

$$\mu = \rho_{ST} - \frac{\rho_t N_1}{N_{ST}} = \frac{\partial TC_{ST}}{\partial N_{ST}} - AC_{ST} - \frac{\rho_t N_1}{N_{ST}} \quad (\text{A.18})$$

The average marginal social cost is $MSC_{ST} = \frac{\partial TC_{ST}}{\partial N_{ST}}$. Therefore, Eq. (A.18) can also be expressed

as: $\mu = MSC_{ST} - AC_{ST} - \frac{\rho_t N_1}{N_{ST}} = MEC_{ST} - \frac{\rho_t N_1}{N_{ST}}$, where MEC_{ST} is the average marginal external cost.

The optimal flat part of toll μ is set such that price equals the average marginal social cost. The optimal demand can be obtained from $D(N_{ST}) = MSC_{ST}$.

A.6 The results in Case II for the single-step toll

As given in Long et al. (2022), the expected social cost in Case II is:

$$TC_{ST}(N_{ST}) = (N_{ST})^2 \zeta(\vec{\phi}, \vec{\phi}) \quad (\text{A.19})$$

where $\zeta(\vec{\phi}, \vec{\phi}) = \beta \vec{\phi} - X(\vec{\phi}, \vec{\phi}) \left[1 - \beta(\vec{\phi} - \vec{\phi})Y(\vec{\phi}) + \frac{2(\beta \vec{\phi} - X(\vec{\phi}, \vec{\phi}))}{(\alpha + \gamma)\vec{\phi}} - X(\vec{\phi}, \vec{\phi})Y(\vec{\phi}) \right]$, $X(\vec{\phi}, \vec{\phi}) =$

$\frac{\beta \vec{\phi} - (\alpha + \gamma)(H(\vec{\phi}) - \vec{\phi})}{(\alpha + \gamma)[H(\vec{\phi})Y(\vec{\phi}) - H(\vec{\phi})/W(\vec{\phi})] + (\alpha + \beta + \gamma)[\vec{\phi}/W(\vec{\phi}) - \vec{\phi}Y(\vec{\phi})] + 1}$. As defined in Long et al. (2022), $\vec{\phi}$ and $\vec{\phi}$ are the

reciprocals of average departure rates during the period from the departure time of the first commuter to the departure time of the first commuter who pays the toll and during the period from the departure time of the first commuter who pays the toll to the end of the tolling period, respectively. $\vec{\phi}$ and $\vec{\phi}$ are

obtained by $\frac{\partial \zeta(\vec{\phi}, \vec{\phi})}{\partial \vec{\phi}} = 0$ and $\frac{\partial \zeta(\vec{\phi}, \vec{\phi})}{\partial \vec{\phi}} = 0$. The total toll revenue follows Eq. (31), where $\rho_t =$

$N_{ST}X(\vec{\phi}, \vec{\phi})$ and $N_1 = N_{ST} - \frac{\rho_t}{w(\vec{\phi})}$ in Case II. As proved in Proposition 20 by Long et al. (2022), $\vec{\phi}$ and

$\vec{\phi}$ are independent of the number of users, and TC_{ST} and ρ_t depend on the travel demand. Hence, the average MEC in Case II is:

$$MEC_{ST} = N_{ST}\zeta(\vec{\phi}, \vec{\phi}) \quad (\text{A.20})$$

The optimal time-invariant part of the toll in Case II is:

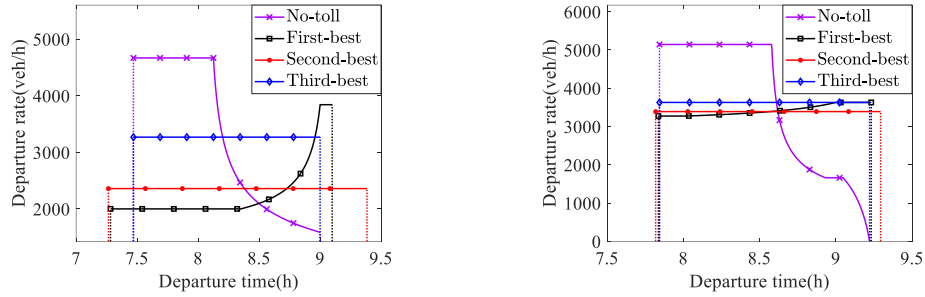
$$\mu = N_{ST} \left[\zeta(\vec{\phi}, \vec{\phi}) - X(\vec{\phi}, \vec{\phi}) \left(1 - \frac{X(\vec{\phi}, \vec{\phi})}{w(\vec{\phi})} \right) \right] \quad (\text{A.21})$$

where $\mu > 0$ in Case II. The generalised price in Case II is:

$$P_{ST} = 2N_{ST}\zeta(\vec{\phi}, \vec{\phi}) \quad (\text{A.22})$$

From Eq. (A.22), the optimal demand under a step toll can be found by $P_{ST} = D(N_{ST})$.

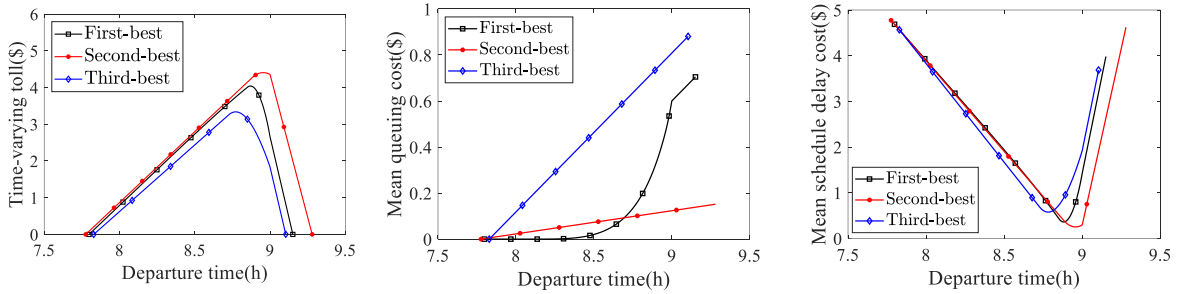
Appendix B. The further results from the numerical study



(a) High uncertainty: $\phi_{min}=0.2$ s/veh

(b) Very low uncertainty: $\phi_{min}=0.9$ s/veh

Fig. B.1. Comparisons of equilibrium departure rates under different time-varying tolling schemes in stochastic bottleneck model, when $\beta/\alpha = 0.3$.



(a) Toll

(b) Mean queuing cost

(c) Mean schedule delay cost

Fig. B.2. Comparisons of toll, mean queuing cost, and mean schedule delay cost under fixed demand when $\phi_{min}=0.8$ s/veh.

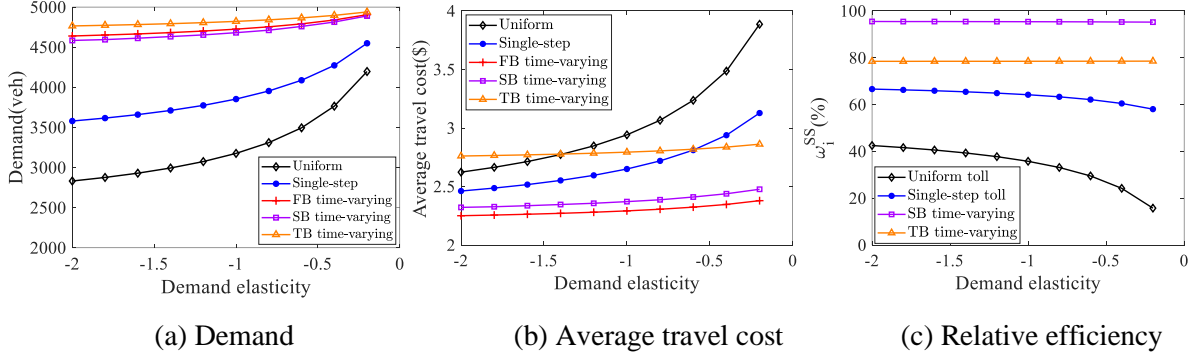


Fig. B.3. The effect of elasticity on the outcomes under price-sensitive demand when $\phi_{min} = 0.75$ s/veh.

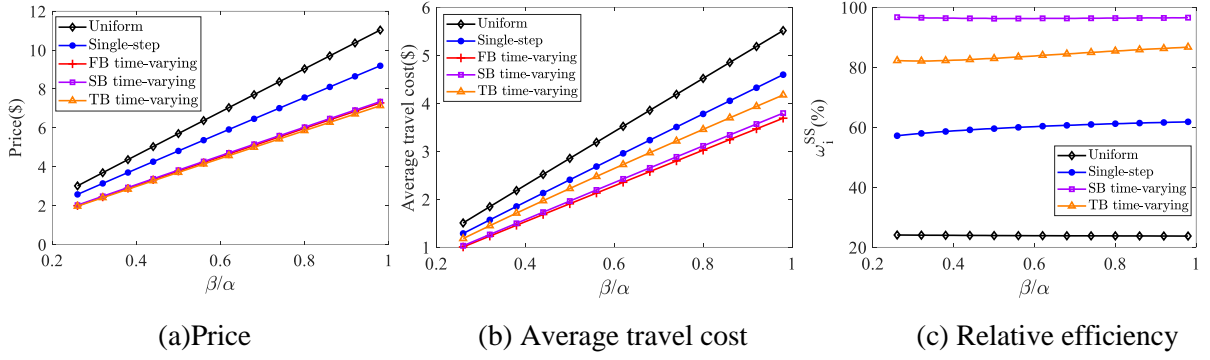


Fig. B.4. The effect of β/α on the outcomes under price-sensitive demand when $\phi_{min} = 0.8$ s/veh.