

## MIT Open Access Articles

*Congestion tolling — Dollars versus tokens: Within-day dynamics*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Seshadri, Ravi, de Palma, André and Ben-Akiva, Moshe. 2022. "Congestion tolling — Dollars versus tokens: Within-day dynamics." *Transportation Research Part C: Emerging Technologies*, 143.

**As Published:** 10.1016/J.TRC.2022.103836

**Publisher:** Elsevier BV

**Persistent URL:** <https://hdl.handle.net/1721.1/148398>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution 4.0 International license



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Transportation Research Part C

journal homepage: [www.elsevier.com/locate/trc](http://www.elsevier.com/locate/trc)

## Congestion tolling — Dollars versus tokens: Within-day dynamics

Ravi Seshadri <sup>a,\*</sup>, André de Palma <sup>b</sup>, Moshe Ben-Akiva <sup>c</sup><sup>a</sup> Department of Technology, Management and Economics, Technical University of Denmark, DTU, 2800 Kgs. Lyngby, Denmark<sup>b</sup> CY Cergy Paris Université THEMA 33 Boulevard du Port, 95000 Cergy-Pontoise, France<sup>c</sup> Massachusetts Institute of Technology (MIT), Room 1-181, 77 Massachusetts Avenue, Cambridge, MA, 02139, United States of America

### ARTICLE INFO

#### JEL classification:

D61

R41

R48

#### Keywords:

Tolls

Tradable credit schemes

Mobility permits

Congestion

Dynamic models

Efficiency

Equity

### ABSTRACT

Tradable credit schemes (or tolling in tokens) are a form of quantity control, which promise to be an appealing alternative to congestion pricing (or tolling in dollars) owing to considerations of revenue neutrality, equity, reduced infrastructure costs, and political acceptability. The comparative performance of the two instruments under uncertainty in demand and supply has only recently received attention in the transportation setting, despite being widely studied for emission markets. In this paper, we add to this literature by considering a tradable credit scheme in a departure time context wherein users are provided an initial endowment of tokens by the regulator and incur a token charge (determined prior to all departures) to travel in a specific time-period. Tokens can be bought and sold within a marketplace at a price determined endogenously by token demand and supply. Two key features of the market model are: (1) the time-of-day dynamics of price is explicitly modeled through a smooth market clearing mechanism in each period, and (2) the selling decisions of users, which determine the distribution of token supply in the market over the day are explicitly modeled. This enables us to study the impact of selling behavior on performance of the credit system. Travel demand is modeled using a logit mixture model and supply consists of static congestion.

Extensive experiments under stochastic demand show that when the tolls (in dollars and tokens) are not day-to-day adaptive, tolling in tokens outperforms tolling in dollars when congestion effects are more severe (e.g., realistic BPR models and steep congestion functions, high demand levels and high day-to-day variability). Importantly, we find that this result is robust with respect to selling behavior in the market, although there can be welfare losses in the quantity control system when selling behavior in the market is excessively irrational. These findings underscore the importance of examining disaggregate market behavior when designing tradable credit schemes. Moreover, when the supply of tokens can be adapted from day to day, the credit system was found to be superior in all tested scenarios, provided the selling behavior of individuals is rational. Finally, even in the case when toll revenues in the price instrument are equally redistributed (often difficult in practice), tolling in tokens (when tokens are equally distributed) is marginally more equitable in scenarios where congestion effects are more severe. These findings make a case for tolling in tokens.

### 1. Introduction

Congestion is a pervasive problem in most transportation networks worldwide and the standard approach to address this issue has been to internalize congestion externalities through a toll in dollars (congestion pricing). Extensive reviews may be found in [de](#)

\* Corresponding author.

E-mail addresses: [ravse@dtu.dk](mailto:ravse@dtu.dk) (R. Seshadri), [andre.de-palma@cyu.fr](mailto:andre.de-palma@cyu.fr) (A. de Palma), [mba@mit.edu](mailto:mba@mit.edu) (M. Ben-Akiva).

<https://doi.org/10.1016/j.trc.2022.103836>

Received 25 October 2021; Received in revised form 1 August 2022; Accepted 3 August 2022

Available online 15 August 2022

0968-090X/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Palma and Lindsey (2011) and Santos and Verhoef (2011). Pricing, however, has long been beset by issues of inequity, complexity, high infrastructure costs and public and political acceptability, notwithstanding the redistribution of toll revenues, whose benefits may take years to be realized (see also Jaensirisak et al. (2005) and de Palma and Lindsey (2020) for more on the acceptability of road pricing). In contrast, tradable credit schemes (or tolling in tokens; also called tradable permit schemes) are a form of quantity control typically characterized by the following features (Fan and Jiang, 2013): (1) a fixed total number of tokens (mobility credits) or 'quota' is pre-specified by the regulator, (2) an initial endowment allocates or distributes the tokens to a selected population (all individuals may not receive tokens), (3) individuals are allowed to buy and sell tokens in a market, (4) use of the road network requires tokens and can be differentiated by time of day, geography, vehicle type etc., and (5) enforcement is necessary to ensure valid trading/consumption of tokens.

Tradable credit schemes have several potential advantages over pricing. First, they are revenue neutral and hence, may not be faced with similar public opposition, more so if the tokens are handed out for free. Second, they are viewed as being less vertically inequitable than pricing (de Palma and Lindsey, 2020). Since the number of tokens each user receives may differ, any regressive effect (very well documented for congestion pricing), which may trigger political opposition, will not occur with tokens. In other words, lower income users who tend to travel less by car can obtain monetary gains by selling their excess credits. Third, implementation costs may be low given developments in information and communications technology. Finally, they provide the ability to directly control quantity, which may be beneficial in some situations, for example, when the price elasticity of demand in the short or medium term is low. Note that the distribution of tokens allows the regulator to cap the maximum number of travelers. This is a pure version of quantity control, which appears to be more flexible than conventional policies (such as the odd and even license plate policy, or reduced speed) to reduce environmental costs. Other forms of quantity control include ramp metering, perimeter control, and metering-based dynamic priority schemes (Lamotte et al., 2022).

Despite the large body of literature on Tradable Credit Schemes (TCS) in the transportation context, relatively little attention has been paid to the comparison of the price and quantity control instruments under uncertainty. In this paper, we consider a tradable credit scheme in a departure time context wherein users are provided an initial endowment of tokens by the regulator and incur a token charge (determined prior to all departures) to travel in a specific time period. Tokens can be bought and sold within a marketplace at a price determined endogenously by token demand and supply. Travel demand is modeled using a logit mixture model and supply consists of static congestion specific to each time period.

This paper contributes to the existing literature on tradable credit schemes and the comparative performance of price and quantity control instruments under uncertainty in several respects. First, from a methodological standpoint, we propose a smooth market clearing mechanism that can be used to model price dynamics across time intervals for a tradable credit scheme (wherein tokens are bankable across time intervals) using a tractable equilibrium approach. Selling decisions of users are explicitly modeled and determine the distribution of token supply in the market over the time of day (to the best of our knowledge, this has yet to be addressed in the literature). This enables us to study the impact of selling behavior on performance of the credit system. Although we use a logit model for convenience, in principle, any continuous model may be used. The proposed framework can also be used to incorporate more complex models of selling behavior (see for example, Dogterom et al. (2017)).

Second, in terms of numerical findings, we obtain new insights into the impact of selling behavior on the performance of a tradable credit scheme relative to pricing under uncertainty: (a) Welfare losses are present in a TCS system with irrational sellers, which has the effect of equalizing token supply across time intervals leading to a deterioration in the performance of the quantity instrument. However, despite these losses, the TCS with fixed token supply is still superior to pricing when congestion effects are severe; (b) In the case when token supply in the TCS can be adapted from day to day, the quantity control system is still not consistently superior to price control as one may expect (it can be inferior in cases with irrational selling behavior when congestion effects are less severe); (c) Even when toll revenues in the price instrument are equally redistributed, tolling in tokens (when tokens are equally distributed) is marginally more equitable in scenarios where congestion effects are more severe. These findings underscore the importance of examining disaggregate market behavior when designing tradable credit schemes.

Finally, we show that the under uncertainty, in a more complex setting involving within-day dynamics (with time dependent tolls, market prices and congestion), quantity control is superior to price control in cases where congestion effects are more severe (steep BPR functions, high demand, high day-to-day variability). The optimal network usage is relatively similar across states for quantity control whereas the optimal toll in dollar amounts varies significantly across states. In this regard, we generalize the findings of de Palma et al. (2018), de Palma and Lindsey (2020), showing that their results on the efficiency differences between the two instruments are robust (holding even in the presence of irrational sellers) and generalize to this more complex setting. These findings make a potential case for tolling in tokens.

The remainder of the paper is organized as follows. Section 2 reviews the existing literature on tradable credits and describes our contributions. Section 3 describes the basic model of supply, demand and equilibrium in the case of deterministic demand (and supply) for both instruments (dollars and tokens). The comparison of the two instruments in this case can be performed analytically. Following this, Section 4 describes the model for stochastic demand for which, the comparisons must be performed numerically. The numerical experiments and findings are described in Section 5. Finally, Section 6 provides concluding remarks and directions for future research.

## 2. Review of literature

Although discussions on the use of tradable credits in the transportation sector date back to Raux (2004), Verhoef et al. (1997) and Goddard (1997), they have received significant attention only in the recent past (detailed reviews may be found in Fan and

Jiang (2013) and Grant-Muller and Xu (2014)). In particular, a variety of tradable credit schemes have been proposed (utilizing largely network and market equilibrium approaches) in the context of mobility management (at the network level) and bottleneck management, and these are discussed in turn.

In the context of mobility management, Yang and Wang (2011) propose a scheme wherein the social planner initially distributes a certain number of credits to all potential travelers, charges a link-specific number of credits for a given link, and allows trading of the credits among travelers. Supply is modeled using static congestion (separable link performance functions) and travelers are assumed to be homogeneous. They demonstrate that for a given set of tolls in tokens (or credit rates) in a general network, the user equilibrium (UE) link flow pattern is unique under standard assumptions, and the credit price at the market equilibrium is unique under some relatively mild additional assumptions (i.e., if all equilibrium path flow patterns contain at least two paths with different credit charges connecting the same O–D pair). The proposed network equilibrium formulation is a variant of the standard UE model with the additional network-wide credit feasibility constraint, which simply states that the total consumption of tokens at equilibrium is less than or equal to the total credit endowment. Extensions that consider heterogeneity in the value of time and multiple user classes are proposed in Zhu et al. (2015) and Wang et al. (2012), whereas He et al. (2013) consider allocations of credits to not just individual travelers, but to transportation firms such as logistics companies and transit agencies. In a similar vein, Pareto-improving credit based congestion management schemes on a general two mode network are investigated in Liu and Nie (2017). Multi-period tradable credit schemes have also been proposed where travelers can transfer credits to future periods within a planning horizon (Miralinaghi and Peeta, 2016, 2019).

In contrast with most studies which adopt an equilibrium perspective, Ye and Yang (2013) examine the price and flow dynamics under a tradable credit scheme using a continuous dynamic model within a finite time horizon, deriving conditions for asymptotic stability and convergence. Recognizing the practical challenges of optimally designing tradable credit schemes, which require knowledge of both demand and supply functions, Wang and Yang (2012), Wang et al. (2014c) propose trial and error methods (when demand functions are not known) for a single road and general networks, respectively. Both the credit tariffs and credit allocation are updated in an iterative fashion based on observed link flows and revealed credit prices.

Nie (2012) examines the effect of transaction costs in a tradable credit scheme for two types of markets: an auction market in which users purchase all of the needed mobility credits through a competitive bidding process (they can also purchase directly from the government at a pre-specified price), and a negotiated market in which users initially receive certain amount of mobility credits from the government and trade with each other through negotiation to meet their needs. A brokerage service is built into both markets to facilitate transactions and accordingly, the users have to pay a commission fee proportional to the value of the trade. The modified UE formulation of Yang and Wang (2011) is extended to incorporate transaction costs for both the auction and negotiated markets. Based on numerical experiments on a toy network, it is shown that an auction market can achieve an equilibrium allocation of mobility credits if the government sets the price suitably (price at which users purchase tokens directly from the government) and the unit transaction cost is lower than the price that the market would reach in the absence of transaction costs. The work also highlights the fact that in case of the negotiated market, the initial allocation of mobility credits may affect the final equilibrium when marginal transaction costs are constant.

The design of tradable credit schemes for mobility management has also been explored using bi-level optimization formulations (mathematical programming problems with equilibrium constraints). For instance, Wang et al. (2014b,a) formulate the continuous network design problem with a tradeable credit scheme as a bi-level programming problem, where the decision variables for the upper level problem are capacity enhancements for selected links whereas the lower level problem determines equilibrium link flows and the credit price. Along similar lines, Wu et al. (2012) proposed a framework that considers decisions of mode/route choice and trip generation on a multimodal transportation network to design efficient and equitable congestion pricing and tradable credit schemes (considering a measure including both net social benefit and equity). They find that the Pareto frontier (with respect to the two aforementioned objectives) of the credit scheme strictly dominates that of congestion pricing although the two schemes achieve the same level of maximum net benefits. Further, their results suggest that tradable credit schemes can be progressive whereas congestion pricing schemes are largely regressive for the tested network. Finally, the literature on mobility management also includes a series of studies on credit-based congestion pricing (CBCP), where credits in CBCPs are allowances used to pay tolls (Kalmanje and Kockelman, 2004; Kockelman and Kalmanje, 2005). The studies involved the use of destination, mode, and departure time choice models to examine the potential impacts of using a CBCP scheme.

Tradable mobility credit schemes have also been proposed to manage bottleneck congestion and achieve peak spreading in an efficient manner. Nie and Yin (2013) developed an analytical framework to model a tradable credit scheme that manages commuters' travel choices in a simple transportation system consisting of two parallel routes. The scheme attempts to persuade commuters to spread their departure times evenly within the rush hour and between primary and alternative routes to mitigate traffic congestion. It defines a 'peak-time' window within which users are charged mobility credits to travel on the primary route and those that avoid either the peak-time window or the primary route may be rewarded with credits (see also Nie, 2015). Tian et al. (2013) investigate the efficiency of a tradable travel credit scheme for managing bottleneck congestion and modal split in a competitive highway/transit network with a continuously distributed value of time. They propose a tradable credit scheme which emulates bottleneck congestion pricing and transit subsidy in a revenue-neutral manner and demonstrate that both the modal split and credit charge at equilibrium are unique.

Xiao et al. (2013) examined the efficiency of a tradable credit system in managing morning commute congestion with both homogeneous and heterogeneous users. They show that even in the presence of heterogeneity, an optimal credit scheme that eliminates the bottleneck queue always exists under the assumption that late arrival is not allowed. More recently, Akamatsu and Wada (2017) proposed a tradable bottleneck credit scheme where the regulator issues link- and time-specific credits permitting

passage through a certain link or bottleneck in a pre-specified time period. They develop a model to describe time-dependent flow patterns at equilibrium under a system of tradable bottleneck permits for general networks and show that the equilibrium obtained under this system is efficient in that it minimizes the social transportation cost. Bao et al. (2019) show that the equilibrium with a tradable credit scheme may not be unique for particular models of traffic congestion, including the first-best solution for the conventional Vickrey's bottleneck model. Liu et al. (2020) examine properties of a TCS in a departure time context using a day-to-day modeling framework and a trip-based MFD model of traffic congestion, and Chen et al. (2021) examine market design aspects of the TCS, including allocation and expiration of tokens, price adjustment, regulator intervention, and the role of transaction fees. Finally, Brands et al. (2020) conduct an interesting lab-in-the-field experiment of tradable credit schemes with virtual mobility behavior and real financial incentives. They adopt a market design, which lets users trade with a price setting intermediary, termed a virtual bank. An incremental price adjustment scheme is adopted and their experiments suggest that it ensures that the price stays largely within the equilibrium range. Overall, their results are promising and indicate that tradable credits can be a viable alternative to pricing in a parking setting.

In contrast, comparisons of price and quantity control under uncertainty in transportation are relatively sparse (for other contexts see Weitzman (1974), Laffont (1977)). Note that in the emissions context, agents are not directly impacted by the externality they generate, while in the transportation context they are (as far as congestion is concerned). For this reason, their results on stochastic demand and stochastic supply are not directly applicable here, even if they are useful as a general guideline. Shirmohammadi et al. (2013) examine the performance of tradable credit systems under demand and supply uncertainty using a toy network. Specifically, they compare the performance of a link differentiated toll system (in dollars) and a mobility credit system that is differentiated by link. However, their analysis does not focus on measures of efficiency but rather examines performance relative to a given target volume of cars. They find strong variations in the credit prices are required to ensure demand matches the specified volume targets.

de Palma et al. (2018) compare the performance (in efficiency terms) of congestion pricing and tradable mobility credit schemes under uncertainty using a simple road network in a stochastic route choice setting (including a public transit alternative). They find that when the tolls (in either dollars or tokens) cannot be adapted from day to day, the credit scheme performs better typically when the slope of the congestion function is steep. Further, when the token supply can be adapted from day to day, the token system always outperforms congestion pricing. More recently, de Palma and Lindsey (2020) ranked the efficiency of tradable permits and tolls for one route, one time period and elastic demand. They study additive and multiplicative demand and cost (capacity) shocks and show that these may lead to qualitatively different results. Their approach is analytical and considers linear and non-linear demand (showing the role of the convexity of demand). They also study the impact of the correlation between demand and supply on the ranking. Along similar lines, Lindsey et al. (2022) compare the allocative efficiency of tradable permits and tolls under uncertainty on a bimodal network and derive a general rule for ranking the efficiency of the two instruments. Rezaeinia et al. (2021) consider the comparison of tolls and permits in a radial network, within the Vickrey framework. Users select mode, departure time and route based on a Nested Logit continuous/discrete framework, and congestion is computed using the METROPOLIS software (de Palma et al., 1997). Tolls and permits are independent of the time of the day. They find that tolls outperform permits if capacities shocks are perfectly correlated. However, the ranking is reversed if capacities shocks are independent.

In summary, literature on the comparative performance of the price and quantity control instruments under uncertainty is relatively sparse. Existing studies (Shirmohammadi et al., 2013; de Palma et al., 2018; de Palma and Lindsey, 2020) are limited to route choice and a single time-period (i.e., they are static models). This paper attempts to extend the literature by generalizing the work of (1) de Palma et al. (2018), who consider static congestion (with parallel routes) and one time period, and (2) de Palma and Lindsey (2020), who consider a single route with elastic demand and static congestion. Here instead, users decide when to travel within a multi-period setting (departure-time choice context), and we propose a smooth market clearing mechanism within each time-period to describe within-day market price dynamics of a TCS when tokens or credits are bankable across time intervals. This enhancement, which involves explicitly modeling selling behavior in the market (to the best of our knowledge, this has not yet been addressed in the literature) is a critical step in moving to large-scale simulation-based assessments of tradable credit schemes on real-world networks. It also enables us to study the impact of selling behavior on performance of the credit system. Given the complexity of our model, comparisons of the two instruments under uncertainty cannot be performed analytically.

Lastly, in comparison with Ye and Yang (2013), two key points are noteworthy. First, they propose a continuous-time evolution model to study price and flow dynamics of a TCS within a finite-length time period of interest (tokens expire at the end of this time period). In contrast, we consider a discrete-time setting with  $N$  time intervals within a day where tokens are bankable across these time intervals. The dynamics of market prices over the day are determined by a market clearing mechanism in each interval. Note that the numerical scheme we utilize to compute the market clearing price in each time interval is largely identical to their price evolution process. However, rather than the evolution process within each time interval, our focus lies in the equilibrium time-of-day prices and their impact on comparative efficiency of the TCS relative to the price instrument. Second, we consider day-to-day variability in the form of demand stochasticity (in contrast with the deterministic demand setting they study).

### 3. Multiperiod model: Deterministic demand

The transportation network of interest consists of a single origin–destination pair connected by a single route (this can be extended to multiple OD pairs and routes with no methodological difficulty). We consider commuting trips performed within a time period  $T$ , which is partitioned into  $I$  sub-periods  $T_1, T_2, \dots, T_I \dots T_I$  so that  $T_1 \cup T_2 \dots \cup T_I = T$ . Note that  $T_1, T_2, \dots, T_I$  may represent any  $I$  time periods within a day, and are not necessarily contiguous (the terms time period and time interval are used interchangeably when referring to  $T_i$ ). We denote the set of the  $I$  time intervals by  $I$ . There are a total of  $N$  users who wish to

travel, and each user performs a single trip or activity during the day in any one of the  $I$  intervals or chooses to stay at home (denoted by  $T_0$ ). A glossary of notation can be found in [Appendix D](#).

We assume that the network is subject to time dependent congestion justifying the need for congestion control in the form of either a price instrument (tolling in dollars) or a quantity instrument (tolling in tokens). Under the price control instrument, users have to pay a toll in dollars  $\tau_i$  to travel in time period  $T_i$  ( $i = 1 \dots I$ ). In case of the quantity control system, we make several assumptions. The regulator distributes a certain number of credits (tokens)  $M$  (known in advance) to each potential user at the beginning of the time period  $T$ . The tokens expire at the end of time period  $T$ , or in other words, their market value is zero at the end of time period  $T$ . Tokens cannot be banked or traded across days (i.e., across periods  $T$ ). This assumption is similar to that of [Brands et al. \(2020\)](#) and is used to mitigate undesirable speculation and hoarding of tokens. Further, the tokens cannot be transferred or used outside the market. Users have to spend a certain number of tokens to travel in time interval  $T_i$ ,  $i = 1 \dots I$ , given by  $\delta_i$  (toll in tokens). Tokens can be bought and sold within a marketplace. The price of the token in time interval  $T_i$  is denoted by  $p_i$ , and is determined endogenously by the demand and supply of tokens in the market in time interval  $T_i$ . In order to avoid speculation, tokens can only be bought if they are needed for travel. This will prevent users from buying tokens for the sole purpose of trying to make a profit in the market, an aspect of the tradable credit scheme that can hamper acceptability. Users must sell all their unused tokens in a single time interval  $T_i$ ,  $i = 1 \dots I$ . Once again, this assumption minimizes excessive transactions for the sole purpose of making a profit. However, it should be pointed out that this assumption can potentially affect market efficiency (see [Brands et al. \(2020\)](#) for a detailed discussion). Nevertheless, we defer relaxation of this assumption to future research. Finally, to improve acceptability and ease of implementation, we assume that all transactions take place at the beginning of each time interval  $T_i$ .

Note that the regulator has the flexibility to institute any desired token allocation scheme including ones wherein users receive an unequal number of tokens. Given that user choices are unaffected by the token allocation (ignoring income effects and transaction costs), this implies that in principle any desired distribution of equity can be achieved through the initial token allocation (see [Yang and Wang \(2011\)](#)).

In this section, we consider the case of deterministic demand (i.e., the number of users  $N$  is deterministic and known). The transportation model is first described, followed by the two instruments (price and quantity) in turn and a comparison with respect to individual benefits, social benefits, and equity.

### 3.1. Transportation model: Demand, supply and equilibrium

The money-metric utility of an individual  $n$  traveling in a time period  $T_i$ ,  $i = 0 \dots I$  ( $i = 0$  denotes the stay at home option) is given by,

$$U^n(T_0) = B_0^n + \mu_n \epsilon_0, \quad (1)$$

$$U^n(T_i) = B_i^n - \alpha^n t_i(X_i) - p_i \delta_i + \mu_n \epsilon_i, \quad i = 1 \dots I, \text{ (quantity control)}$$

$$U^n(T_i) = B_i^n - \alpha^n t_i(X_i) - \tau_i + \mu_n \epsilon_i, \quad i = 1 \dots I, \text{ (price control)},$$

where  $X_i$  is the flow in time period  $T_i$ ,  $\epsilon_i$  is an i.i.d. Gumbel disturbance term and  $\mu_n, B_i^n, \alpha^n$  are individual specific parameters with  $\log(\mu_n), B_i^n, \log(\alpha^n)$  normally distributed. The alternative specific benefit  $B_i^n$  incorporates time-period specific scheduling preferences, or alternatively, time-period specific schedule delay costs.  $\alpha^n$  and  $\mu_n$  are the value of time and scale parameter, respectively, for individual  $n$ .

A standard BPR type function is assumed to model the travel time  $t_i(X_i)$  in time period  $i$ ,

$$t_i(X_i) = t_i^{FF} \left( 1 + \alpha_i (X_i/C_i)^{\beta_i} \right), \quad (2)$$

where  $t_i^{FF}$  is the free flow time in period  $i$ ,  $C_i$  is a capacity associated with time period  $i$  and  $\alpha_i, \beta_i$  are function parameters. Note that we assume that the  $I$  time periods are independent in the sense that congestion does not spill over from one period to the next (in other words we have  $I$  static models). This is a *time-sliced static traffic assignment* common in planning applications ([Nakayama and Connors, 2014](#)), which admittedly may not realistically capture the dynamics of congestion and flow propagation. However, it suffices for the comparison of the instruments in our context. A more detailed treatment of the dynamics of congestion is beyond the scope of the paper although we note that the methodology proposed is generic and not restricted to any specific type of congestion model.

Further, let  $\theta_n = (\mu_n, B_i^n, \alpha^n)$  denote the vector of parameters for individual  $n$ . The probability of individual  $n$  choosing to travel in time period  $T_i$  is given by,

$$p(T_i | \theta_n) = \frac{\exp[V^n(T_i, \theta_n)]}{\sum_{j=0 \dots I} \exp[V^n(T_j, \theta_n)]}, \quad (3)$$

where  $V^n(T_i, \theta_n)$  is given by,

$$V^n(T_0, \theta_n) = (1/\mu_n) (B_0^n), \quad (4)$$

$$V^n(T_i, \theta_n) = (1/\mu_n) (B_i^n - \alpha^n t_i(X_i) - p_i \delta_i), \quad i = 1 \dots I, \text{ (quantity)}$$



$$V^n(T_i, \theta_n) = (1/\mu_n) (B_i^n - \alpha^n t_i(X_i) - \tau_i), \quad i = 1 \dots I, \text{ (price).}$$

The number of travelers traveling in the  $I$  periods  $X_i$  ( $i = 1 \dots I$ ), given  $\tau_i$  ( $i = 1 \dots I$ ) in the price system and  $p_i \delta_i$  ( $i = 1 \dots I$ ) in the quantity system, are obtained by solving the fixed-point problem (note that  $X_0 = N - \sum_{i=1}^I X_i$ ),

$$X_i = \sum_{n=1}^{n=N} \frac{\exp[V^n(T_i, \theta_n)]}{\sum_j \exp[V^n(T_j, \theta_n)]}, i = 1 \dots I. \tag{5}$$

Since the set of demand feasible flows  $D = (X : \sum_{i=0 \dots I} X_i = N; X_i \geq 0, i = 0 \dots I)$  forms a closed and convex set, and the right hand side of Eq. (5) is a continuous function of flows, Brouwer’s fixed-point theorem implies that a solution exists to the fixed-point problem in Eq. (5). On the conditions for solution uniqueness we refer the reader to Konishi (2004) and Lindsey (2004).

### 3.2. Tolls in dollars: Price control

In the price control system, the regulator is assumed to have knowledge of the demand  $N$  and sets the tolls in dollars  $\tau_i$  ( $i = 1 \dots I$ ) to maximize total welfare (defined as the sum of consumer surplus and regulator revenue). Consumer surplus is calculated as the *logsum* measure (see Small and Rosen (1981) and De Jong et al. (2007)) and regulator revenue is simply the total toll revenue. In our context, the definition of welfare is straightforward since there are no income effects (Anderson et al., 1992). Thus, determination of the tolls can be formulated as the following optimization problem,

$$\text{Max}_{\tau_1, \tau_2, \dots, \tau_I} \Omega_p = \sum_{n=1}^N \mu_n \log \left( \sum_{j=0 \dots I} \exp[V^n(T_j, \theta_n)] \right) + \sum_{j=1 \dots I} \tau_j X_j \tag{6}$$

s.t

$$X_i = \sum_{n=1}^N \frac{\exp[V^n(T_i, \theta_n)]}{\sum_j \exp[V^n(T_j, \theta_n)]}, i = 1 \dots I, \text{ with}$$

$$\sum_{i=0 \dots I} X_i = N, \quad X_i \geq 0, i = 0 \dots I.$$

The optimum welfare obtained by solving (6) above and the corresponding optimum welfare and tolls in dollars are denoted by  $\Omega_p^*$  and  $\tau^* = (\tau_i^*, i = 1 \dots I)$  respectively. Note that the optimization problem in Eq. (6) is non-convex due to the non-linear equilibrium constraints. Solution approaches for this problem are discussed in Section 5.1.

### 3.3. Tolls in tokens: Quantity control

Recall that in the quantity control system, the regulator distributes a certain number of credits (tokens)  $M$  to each potential user at the beginning of a time period  $T$ , and the number of tokens required to travel in time interval  $T_i, i = 1 \dots I$  is given by  $\delta_i$ .

#### 3.3.1. Demand for tokens

For a given set of tolls in tokens and market prices  $\delta_i, p_i$  ( $i = 1 \dots I$ ), the total demand for tokens in period  $T_i$  is  $X_i \delta_i$  and the total amount of tokens possessed by people traveling in period  $T_i$  is  $X_i M$ , where  $M$  is the initial token endowment. Thus, the demand for tokens in period  $T_i$  ( $i = 1 \dots I$ ), is given by,

$$D_i = X_i \text{Max}(0, (\delta_i - M)). \tag{7}$$

where  $X_i$  ( $i = 1 \dots I$ ) are obtained from the solution to Eq. (5). Note that in the above we assume that the token endowment to each user is equal ( $M$ ). This assumption can be relaxed.

#### 3.3.2. Supply of tokens

We assume that the decision to sell tokens is made after the mobility decision and hence, users sell all unused tokens. Note also that it is assumed that there is some scarcity in the system, namely that the number of tokens available is less than what would be consumed if the tokens were free. Formally, if  $\bar{X}_i$  ( $i = 0 \dots I$ ) denotes the equilibrium flows in the absence of tolls, we assume that  $\sum_{i=1 \dots I} \bar{X}_i \delta_i > M N$ .

In the case that there is no congestion (i.e.,  $t_i$  is constant), we have  $\delta_i = 0$  ( $i = 1 \dots I$ ), or equivalently,  $\delta_i < M$  ( $i = 1 \dots I$ ), and hence, the price of tokens  $p_i$  ( $i = 1 \dots I$ ) is zero and the tokens have no effect/value. We focus on the case where congestion effects are present, and hence, it should be the case that at least one of  $\delta_1, \delta_2, \dots, \delta_I$  is larger than  $M$  (note that  $\delta_0 = 0$ ), which implies that  $p_i > 0$ , for at least one period  $T_i$ . Let the subset of time periods where  $\delta_i > M$  be denoted by  $\tilde{I}$ , then the subset of time periods where  $\delta_i < M$  is  $I \setminus \tilde{I}$ .

As noted before, we assume that tokens can only be bought if they are needed for travel. Further, we assume that the regulator requires the user to sell all her unused tokens in one time interval. This assumption is made from a practical standpoint to ensure simplicity of the system. Since the tokens are worthless at the end of the last time interval, no user will keep tokens unused.

Observe that sellers are individuals who have chosen to travel in an interval  $T_j, j \in I \setminus \tilde{I}$  whereas buyers are individuals who have chosen to travel in  $T_i, i \in \tilde{I}$ . We assume that the token price in interval  $T_i, i \in \tilde{I}$  perceived by a seller is given by,

$$\bar{p}_i = p_i + \bar{\epsilon}_i, \quad (8)$$

where  $p_i$  is the market clearing price of the token and  $\bar{\epsilon}_i$  is an i.i.d. Gumbel term with scale parameter  $\bar{\mu}$ . The rationale for the additive term is the following. There may exist day-to-day fluctuations around the average equilibrium price (long-run market clearing price that we have referred to as  $p_i$  thus far) for a given interval. These fluctuations are the result of the operation of the market and dynamic adjustments (they may also arise due to the intrinsically stochastic nature of traffic flows and other sources of uncertainty, see for example (Chen et al., 2002)). Travelers do not have exact knowledge of these fluctuations when making the travel decision at the beginning of the day and hence make their decision based on the long-run average prices (Eq. (1)). The selling decisions on the other hand are made at the beginning of each interval, within the day. When the users make their selling decisions, they have additional information about the day including actual prices revealed in the intervals thus far. This could give rise to different perceptions of the actual prices. Thus, there is a distribution of perceived prices across the population (around the long-run average market clearing prices) that determines the selling decisions of users.

Note that Eq. (8) permits an alternative interpretation, i.e.,  $\bar{p}_i$  is the utility obtained from selling in interval  $i$ . The selling decision may be based on a number of factors apart from price that are subsumed into the error term. These could be the result of factors like different reservation prices, lack of attention satisfying behavior, endowment effects, loss aversion, risk aversion, multiple trips during the day (because of which users wish to sell at different intervals) and so on, all of which are not incorporated within the model.

For a seller, we model the choice of a selling interval using a simple logit model based on Eq. (8), where the systematic utility of selling in interval  $T_i, i \in \tilde{I}$  is based on the average market price in the interval, and given by  $\bar{p}_i$ . Consider a user  $n$  who has decided to travel in period  $T_j, j \in I \setminus \tilde{I}$ . We assume that the amount of tokens to sell does not influence the time period of selling. The probability of the user  $n$  selling his/her token in interval  $T_i, i \in \tilde{I}$ ,  $Q_i$  is assumed to be,

$$Q_i = \text{Prob}[\bar{p}_i > \max_{j \neq i; j \in \tilde{I}} \bar{p}_j] = \frac{\exp(p_i/\bar{\mu})}{\sum_{j \in \tilde{I}} \exp(p_j/\bar{\mu})}. \quad (9)$$

The supply of tokens from users traveling in  $T_j, j \in I \setminus \tilde{I}$  who have decided to sell in period  $T_i, i \in \tilde{I}$  is given by,

$$S_i = \sum_{j \in I \setminus \tilde{I}} Q_i X_j (M - \delta_j) + Q_i X_0 M. \quad (10)$$

Or,

$$S_i = Q_i \left( \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M \right) \quad i \in \tilde{I}. \quad (11)$$

The total supply of tokens is given by,

$$S = \sum_{i \in \tilde{I}} S_i = \sum_{i \in \tilde{I}} Q_i \left( \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M \right) = \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M. \quad (12)$$

We next study the market clearing condition.

### 3.3.3. Market clearing

The market clearing conditions in each interval  $T_i, i \in \tilde{I}$  imply,

$$S_i = D_i, \forall i \in \tilde{I}, \quad (13)$$

Or,

$$Q_i \left( \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M \right) = X_i (\delta_i - M), \forall i \in \tilde{I}. \quad (14)$$

Further, the total demand for tokens is given by,

$$D = \sum_{i \in \tilde{I}} D_i = \sum_{i \in \tilde{I}} X_i (\delta_i - M). \quad (15)$$

Thus, market clearing requires  $D = S$ , or,

$$\begin{aligned} \sum_{i \in \tilde{I}} X_i (\delta_i - M) &= \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M, \\ \text{Or, } \sum_i X_i \delta_i &= M \left( \sum_i X_i + X_0 \right) = MN. \end{aligned} \quad (16)$$

The market clearing conditions in all periods is satisfied ( $S_i = D_i, i \in \tilde{I}$ ). However, the demand (or supply) for tokens in different intervals may be different, i.e.,  $D_1 \neq D_2 \neq D_3 \dots$  (or  $S_1 \neq S_2 \neq S_3 \dots$ ).



### 3.3.4. Computation of market clearing prices

For a given vector of tolls in tokens ( $\delta_i, i = 1 \dots I$ ), the market clearing price in each time interval can be computed through the following iterative process, where the price in iteration  $w + 1$  is given by,

$$p_i^{w+1} = p_i^w + h(D_i^w - S_i^w), \quad i = 1 \dots I, \tag{17}$$

where  $h'(\cdot) > 0$  and  $h(0) = 0$ ;  $D_i^w, S_i^w$  denote the demand and supply of tokens in interval  $i$  and iteration  $w$ , respectively. This is a standard cobweb adjustment process. It should be noted that the price adjustment process above is merely a numerical method to compute the market clearing prices and does not imply that the market actually operates in this manner. Otherwise, users could discover this and make use of the knowledge of this process strategically, which will lead to a complex game theoretic problem outside the scope of this paper (and possibly behaviorally unrealistic in any case). Alternatively, a simpler iterative process can be used to compute the market clearing prices ( $\forall i \in I$ ):

$$p_i^{w+1} = \begin{cases} p_i^w + \Delta_p & D_i^w > S_i^w \\ p_i^w & D_i^w = S_i^w \\ p_i^w - \Delta_p & D_i^w < S_i^w, \end{cases} \tag{18}$$

where  $\Delta_p > 0$ . More details on how this price adjustment scheme is used within the solution framework (to compute market clearing prices) are provided in Section 5.1.

**Conjecture 1.** *The equilibrium prices for the  $I$  time intervals satisfy  $\lim_{\tilde{\mu} \rightarrow 0} p_i = p^*, i \in \tilde{I}$  for any  $\delta_1, \delta_2, \dots, \delta_I$ .*

A discussion of this is provided in Appendix C.

### 3.3.5. Optimization

As before, let  $\tilde{I}$  denote the subset of time periods where  $\delta_i > M$ . We assume that the regulator has knowledge of the demand  $N$  and sets the tolls in tokens for time periods  $1, 2 \dots i-1, i+1 \dots I$ , and the supply of tokens  $M$  ( $\delta_i$  is normalized to 1 token without loss of generality; the period  $i$  can be chosen arbitrarily, for example the morning peak hour) that maximizes total welfare. The welfare in the case of quantity control is simply the consumer surplus, since the regulator revenue is zero given that the instrument is revenue neutral. However, the calculation of consumer surplus differs slightly from the price control case. Recall that in the definition of the utility in Eq. (1), the entire token payment for each individual contributes to a real monetary disutility or monetary loss and does not account for the initial free token allocation or the monetary gains for the users who sell tokens. Therefore, this monetary loss has to be added back to the consumer surplus to reflect the free allocation from the regulator and the monetary gains for sellers. Note that monetary transfers between users and between the regulator and users have no impacts on overall welfare, provided the transfers do not alter or change behavior. Thus, the determination of the toll in tokens can be formulated as the following optimization problem,

$$\text{Max}_{\delta_1, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_I, M} \Omega_q = \sum_{n=1}^N \mu_n \log \left( \sum_{j=0 \dots I} \exp[V^n(T_j, \theta_n)] \right) + \sum_{j=1 \dots I} p_j \delta_j X_j \tag{19}$$

s.t

$$X_i = \sum_{n=1}^N \frac{\exp[V^n(T_i, \theta_n)]}{\sum_j \exp[V^n(T_j, \theta_n)]}, \quad i = 1 \dots I, \text{ with}$$

$$\sum_{i=0 \dots I} X_i = N; \quad X_i \geq 0, \quad i = 0 \dots I,$$

where the equilibrium prices  $\mathbf{p} = (p_1, p_2, \dots, p_I)$  satisfy the market equilibrium conditions (and  $Q_i$  is given by Eq. (9)),

$$Q_i \left( \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M \right) = X_i (\delta_i - M), \quad \forall i \in \tilde{I}.$$

The optimum welfare obtained by solving (19) above and the corresponding optimal tolls in tokens are denoted by  $\Omega_q^*$  and  $\delta^* = (\delta_i^*, i = 1 \dots I)$  respectively. The associated market clearing prices are denoted by  $p_i^*, i = 1 \dots I$ .

We refer the reader to Appendix A for a discussion of the existence and uniqueness of a solution for the market and network equilibrium constraints in Eq. (19). Convergence of the price adjustment schemes in Section 3.3.4 is also discussed.

## 3.4. Comparison

In the deterministic case, the comparison of the two instruments is trivial and can be performed analytically. The two instruments, when optimally chosen, yield identical social welfare. This is shown in Proposition 1 below.

**Proposition 1.** *Under deterministic demand and supply, the two instruments, price and quantity, when optimally chosen, are equivalent.*

**Proof.** Refer Appendix B.

#### 4. Multiperiod model: Stochastic demand

In the discussion thus far, transportation demand and supply were assumed to be deterministic. We now turn our attention to the case where the demand is stochastic.

##### 4.1. Transportation model: Demand, supply and equilibrium

Assume that there exist two days or states of nature  $s1$  and  $s2$  (denoted by  $sk, k = 1, 2$ ), where the alternative specific benefit to travel for individual  $n$ , varies across the days, taking values  $B_i^{n,s1}$  for period  $i$  ( $i = 1, 2, \dots, I$ ) with probability  $q$  and values  $B_i^{n,s2}, i = 1, 2, \dots, I$  with probability  $1 - q$ . In other words, the source of day-to-day variability or stochasticity is on the demand side and arises due to fluctuations in the scheduling preferences of travelers (note that the total number of users is fixed). The variability in scheduling preferences may be due to special events, weather etc. leading to a higher number of users who wish to travel during the peak period. Stochasticity may also arise from external factors affecting supply such as incidents and accidents, or factors affecting both demand and supply. The methodological framework can be extended in a straightforward manner to model these cases as well. Note also that the framework can be extended to several states of nature or even for a continuum of states with no methodological difficulty.

The systematic utilities to travel in time period  $T_i$  on day  $sk, k = 1, 2$  (denoted by  $T_i^{sk}$ ) are given by,

$$V_{sk}^n(T_0^{sk}, \theta_n) = (1/\mu_n) (B_0^n), k = 1, 2 \tag{20}$$

$$V_{sk}^n(T_i^{sk}, \theta_n) = (1/\mu_n) (B_i^{n,sk} - \alpha^n \tau_i(X_i^{sk}) - p_i^{sk} \delta_i), i = 1 \dots I, k = 1, 2 \text{ (quantity)}$$

$$V_{sk}^n(T_i^{sk}, \theta_n) = (1/\mu_n) (B_i^{n,sk} - \alpha^n \tau_i(X_i^{sk}) - \tau_i), i = 1 \dots I, k = 1, 2 \text{ (price)},$$

where  $p_i^{sk}$  and  $X_i^{sk}$  are the token price and number of individuals traveling in time period  $T_i^{sk}$ , respectively. Note that in case of the price system, the terms  $p_i^{sk} \delta_i$  are replaced by  $\tau_i$ . As before, for a given set of tolls in tokens ( $\delta_i, i = 1 \dots I$ ) and token prices ( $p_i^{sk}, i = 1 \dots I; k = 1, 2$ ) — or tolls in dollars ( $\tau_i, i = 1 \dots I$ ) in the case of price control—,  $X_i^{sk}$  for  $k = 1, 2$  can be determined by solving the following fixed-point problem (note that  $X_0^{sk} = N - \sum_{i=1}^I X_i^{sk}$ ),

$$X_i^{sk} = \sum_{n=1}^{n=N} \frac{\exp[V_{sk}^n(T_i^{sk}, \theta_n)]}{\sum_j \exp[V_{sk}^n(T_j^{sk}, \theta_n)}, i = 1 \dots I. \tag{21}$$

##### 4.2. Tolls in dollars: Price control

In case of the price instrument, we assume that the regulator may not wish to change the tolls from day to day for reasons of acceptability and ease of implementation (or may not have knowledge of the specific realization of the state of nature). For instance, in the ERP system of Singapore, tolls are revised only once every few months and do not vary from day to day. Thus, in the case of stochastic demand, the regulator sets the tolls in tokens for the  $I$  time periods that maximizes *expected* total welfare, formulated as the following optimization problem,

$$\begin{aligned} \text{Max}_{\tau_1, \tau_2, \dots, \tau_I} \quad & q \left\{ \sum_{n=1}^N \mu_n \log \left( \sum_{j=0 \dots I} \exp[V_{s1}^n(T_j^{s1}, \theta_n)] \right) + \sum_{j=1 \dots I} \tau_j X_j^{s1} \right\} \\ & + (1 - q) \left\{ \sum_{n=1}^N \mu_n \log \left( \sum_{j=0 \dots I} \exp[V_{s2}^n(T_j^{s2}, \theta_n)] \right) + \sum_{j=1 \dots I} \tau_j X_j^{s2} \right\} \end{aligned} \tag{22}$$

s.t

$$X_i^{sk} = \sum_{n=1}^{n=N} \frac{\exp[V_{sk}^n(T_i^{sk}, \theta_n)]}{\sum_j \exp[V_{sk}^n(T_j^{sk}, \theta_n)}, i = 1 \dots I; k = 1, 2, \text{ with}$$

$$\sum_{i=0 \dots I} X_i^{sk} = N, k = 1, 2; \quad X_i^{sk} \geq 0, i = 0 \dots I, k = 1, 2.$$

##### 4.3. Tolls in tokens: Quantity control

We distinguish two configurations of the quantity control system. First, in the case of adaptive token supply, the supply of tokens can vary by day and is denoted  $M^{s1}$  and  $M^{s2}$ , whereas in the case of fixed token supply, it is assumed that the total supply of tokens is fixed across days i.e.,  $M^{s1} = M^{s2} = M$ . From the standpoint of implementation, adapting the token supply is likely to be far easier than adapting the tolls in tokens (or dollars), which may involve communicating a complex tariff structure (in a general network) to commuters.

The market clearing conditions in Eq. (14) now apply in each time interval for both days and are given by (the same notation as before is used with the added superscript  $sk$  to denote the day),

$$Q_i^{sk} \left( \sum_{j \in I \setminus \tilde{I}^{sk}} X_j^{sk} (M^{sk} - \delta_j) + X_0^{sk} M^{sk} \right) = X_i^{sk} (\delta_i - M^{sk}), \forall i \in \tilde{I}^{sk}, k = 1, 2. \tag{23}$$

In the case of fixed token supply, we assume that the regulator does not have knowledge of the specific realization of the state of nature (or day) and hence, sets the tolls in tokens for time periods  $1, 2 \dots i - 1, i + 1 \dots I$ , and the supply of tokens  $M$  (as before  $\delta_i$  is normalized to 1 token without loss of generality) that maximizes *expected* total welfare, formulated as the following optimization problem,

$$\begin{aligned} \text{Max}_{\delta_1, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_I, M} \quad & q \left\{ \sum_{n=1}^N \mu_n \log \left( \sum_j \exp[V_{s1}^n(T_j^{s1}, \theta_n)] \right) + \sum_j p_j^{s1} \delta_j X_j^{s1} \right\} \\ & + (1 - q) \left\{ \sum_{n=1}^N \mu_n \log \left( \sum_j \exp[V_{s2}^n(T_j^{s2}, \theta_n)] \right) + \sum_j p_j^{s2} \delta_j X_j^{s2} \right\} \end{aligned} \tag{24}$$

s.t

$$\begin{aligned} X_i^{sk} &= \sum_{n=1}^{n=N} \frac{\exp[V_{sk}^n(T_i^{sk}, \theta_n)]}{\sum_j \exp[V_{sk}^n(T_j^{sk}, \theta_n)}, i = 1 \dots I; k = 1, 2, \text{ with} \\ \sum_{i=0 \dots I} X_i^{sk} &= N, k = 1, 2; \quad X_i^{sk} \geq 0, i = 0 \dots I, k = 1, 2, \end{aligned}$$

where the equilibrium prices  $\mathbf{p}^{sk} = (p_1^{sk}, p_2^{sk}, \dots, p_I^{sk}), k = 1, 2$  satisfy the market equilibrium conditions:

$$Q_i^{sk} \left( \sum_{j \in I \setminus \tilde{I}^{sk}} X_j^{sk} (M - \delta_j) + X_0^{sk} M \right) = X_i^{sk} (\delta_i - M), \forall i \in \tilde{I}^{sk}, k = 1, 2.$$

In the case of adaptive token supply, we assume that the regulator has knowledge of the specific realization of the state of nature (or day) and sets the tolls in tokens for time periods  $1, 2 \dots i - 1, i + 1 \dots I$ , and the supply of tokens  $M^{s1}, M^{s2}$  (as before  $\delta_i$  is normalized to 1 token without loss of generality) that maximizes *expected* total welfare, formulated as the following optimization problem,

$$\begin{aligned} \text{Max}_{\delta_1, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_I, M^{s1}, M^{s2}} \quad & q \left\{ \sum_{n=1}^N \mu_n \log \left( \sum_j \exp[V_{s1}^n(T_j^{s1}, \theta_n)] \right) + \sum_j p_j^{s1} \delta_j X_j^{s1} \right\} \\ & + (1 - q) \left\{ \sum_{n=1}^N \mu_n \log \left( \sum_j \exp[V_{s2}^n(T_j^{s2}, \theta_n)] \right) + \sum_j p_j^{s2} \delta_j X_j^{s2} \right\} \end{aligned} \tag{25}$$

s.t

$$\begin{aligned} X_i^{sk} &= \sum_{n=1}^{n=N} \frac{\exp[V_{sk}^n(T_i^{sk}, \theta_n)]}{\sum_j \exp[V_{sk}^n(T_j^{sk}, \theta_n)}, i = 1 \dots I; k = 1, 2, \text{ with} \\ \sum_{i=0 \dots I} X_i^{sk} &= N, k = 1, 2; \quad X_i^{sk} \geq 0, i = 0 \dots I, k = 1, 2, \end{aligned}$$

where the equilibrium prices  $\mathbf{p}^{sk} = (p_1^{sk}, p_2^{sk}, \dots, p_I^{sk}), k = 1, 2$  satisfy the market equilibrium conditions,

$$\Rightarrow Q_i^{sk} \left( \sum_{j \in I \setminus \tilde{I}^{sk}} X_j^{sk} (M^{sk} - \delta_j) + X_0^{sk} M^{sk} \right) = X_i^{sk} (\delta_i - M^{sk}), \forall i \in \tilde{I}^{sk}, k = 1, 2.$$

#### 4.4. Comparison

In contrast with the deterministic case, when demand (or supply) is stochastic, the comparison of the price and quantity control instruments cannot be performed analytically. Hence, we perform the comparison numerically.

### 5. Numerical experiments: Stochastic demand

#### 5.1. Experimental design

The two instruments are compared using a synthetic example across a wide range of demand and supply inputs. The setting we consider involves three time intervals (e.g., early morning, peak, off-peak), i.e.,  $I = 3$ , wherein there are  $N$  potential travelers who may choose to either travel in one of three time periods ( $T_i, i = 1 \dots 3$ ) or cancel their trip (option  $T_0$ ). The stochasticity or

**Table 1**  
Fixed factors.

Parameter	Time period		
	$T_1$	$T_2$	$T_3$
$B_i^{n,s1}$ [Mean] (\$)	7.5	10.0	7.5
$B_i^{n,s1}$ [SD] (\$)	0.25	0.25	0.25
Free flow time (min)	13	13	13
Capacity (vehicles/time period)	350	350	350
$\alpha$ (BPR parameter)	0.175	0.15	0.2

**Table 2**  
Variable factors and levels.

Factors	Levels			
COV of mobility model scale $\mu_n$	0.0	0.2	0.33	0.5
Scale of selling model ( $\bar{\mu}$ )	1	1.5	2	100000
Number of travelers ( $N$ )	1400	1550	1700	
BPR Congestion coefficient ( $\beta$ )	3	4	6	
Benefit Difference in \$ ( $\Delta$ )	3	4	5	

variability in demand – as noted in Section 4 – is modeled by varying the levels of the alternative specific benefit to travel (in time periods) on the two days,  $B_i^{n,sk}, i = 1 \dots 3, k = 1, 2$  (refer Eq. (20)). Thus, the source of day-to-day variability or stochasticity is on the demand side and arises due to fluctuations in the scheduling preferences of travelers, which may arise due to special events, weather etc. leading to a higher number of users who wish to travel. The mean and standard deviation of the alternative specific benefit to cancel trip ( $B_0^{n,sk}, k = 1, 2$ ) are normalized to zero, and the probability  $q$  is assumed to be 0.5.

The values of the fixed factors are shown in Table 1. The capacities are set based on the range of demand values (varies with scenario, see Table 2) to yield a ratio of congested to free flow travel time (in the absence of tolls) in the range 1.25–2.5. The free flow travel time is set to be 13 minutes (assuming a free flow speed of 60 km/hr, this corresponds to a trip length of 13 km, which is in the range of average trip lengths in typical urban transportation networks). The mean of the alternative specific benefit is assumed to be higher in period two to represent peaking effects and commute behavior (note that the table describes the distribution of the alternative specific benefits in the different time periods for day  $s1$ ; the values on day  $s2$  vary with the scenario, and are part of the experimental design, which is described later in the section). The coefficient of variation of the alternative specific benefit is assumed to be lower in period  $T_2$  reflecting a morning commute context where work start times are largely in this time interval. Further, we introduce some asymmetry in periods  $T_1$  and  $T_3$  through the BPR congestion function, which could potentially reflect choices of different routes in these periods. The mean and standard deviation of the value of time are assumed to be 0.33\$ per min (around 20\$ per hr) and 0.067\$ per min (around 4\$ per hr) respectively (refer Prato et al. (2014), Hess et al. (2005), Cirillo and Axhausen (2006) for empirical evidence; note that the literature reports a wide range of values for the coefficient of variation, we adopt a conservative value of 0.2).

In the experimental design, five factors are varied, which include the coefficient of variation (COV) of the scale parameter  $\mu_n$  in the mobility model (the mean of  $\mu_n$  is fixed at 1.5), the scale parameter of the selling model  $\bar{\mu}$ , total number of users  $N$ , the congestion coefficient  $\beta$ , and the benefit difference between the two days  $\Delta$ . The factor levels are shown in Table 2. A total of 432 test instances or scenarios ( $4^2 \times 3^3$ ) were simulated.

Several additional points are noteworthy. First, in all the scenarios wherein  $\mu_n$  is deterministic (in other words, COV of  $\mu_n$  is zero), the standard deviations of all other randomly distributed parameters (i.e.,  $\alpha^n; B_i^{n,sk}, i = 1 \dots 3, k = 1, 2$ ) are also set to zero. Thus, this subset of scenarios represents the setting with no heterogeneity in the mobility model. Second, in order to set the values of the alternative specific benefits for a given scenario with a benefit difference  $\Delta$ , the values of  $B_i^{n,s1}, i = 1 \dots 3$ , are first sampled (based on the mean and standard deviation in Table 1), and  $B_i^{n,s2}, i = 1 \dots 3$  is given by,  $B_i^{n,s2} = B_i^{n,s1} + \Delta, i = 1 \dots 3$ . Third, the scenarios with  $\bar{\mu} = 100000$  represent non-rational market behavior or a purely random selling model (i.e.,  $\bar{\mu} \rightarrow \infty$ ).

The two instruments (tolling in dollars and tolling in tokens) are compared across the 432 test instances based on the optimum social welfare obtained by solving the optimization problems in Eqs. (22), (24) and (25). These are solved as bi-level problems; given the non-convexity of the problems, two approaches are adopted for solving the upper-level problem.

For price control (Eq. (22)), in the first approach, the optimization problem at the upper-level (decision variable is the toll in dollars;  $\tau_i, i = 1 \dots I$ ) is solved using a genetic algorithm (GA). The GA is well suited to mathematical programs with equilibrium constraints (MPEC) (Zhang and Yang, 2004). A population size of 50 with a maximum of 300 generations and a crossover fraction of 0.8 is used (for more details see Deep et al. (2009)). This is implemented directly using the MATLAB *ga* routine. For each evaluation of the objective function for a given candidate solution of the toll in dollars  $\tau_i, i = 1 \dots I$ , the lower-level Stochastic User Equilibrium (SUE) problem is solved as a system of non-linear equations using the *lsqnonlin* routine to determine the flows ( $X_i^{sk}, i = 0 \dots I, k = 1, 2$ ). In the second approach, the GA is replaced by the sequential quadratic programming algorithm, which has also been applied to non-convex MPEC problems (Meng et al., 2004). However, in this case, given the non-convexity, 25 randomly generated starting points are used for the optimization algorithm (the value of 25 was arrived at empirically based on preliminary experiments wherein it was found that increasing the number of starting points beyond 25 did not yield improvements in the objective value). This is implemented using the *fmincon* routine in MATLAB.

**Table 3**  
Summary statistics: Welfare differences.

Statistic	Welfare difference (\$)					Percentage diff.		Welfare (\$)	
	ADP-NT	SP-NT	SQ-NT	SQ-SP	ADP-SP	$\frac{(SQ-SP)}{NT}$	$\frac{(SQ-SP)}{ADP}$	NT	ADP
Mean	1776	1555	1681	127	222	2.0	1.4	7347	9123
Median	1569	1416	1501	90	169	1.2	1.0	7443	9132
Min	575	517	104	-545	39	-6.8	-6.2	4645	8024
Max	4667	3890	4590	701	778	13.9	7.2	9021	10336
25th per.	1003	867	874	15	103	0.2	0.2	6613	8668
75 per.	2181	1896	2076	213	286	3.3	2.4	8131	9540
$N_S(>0)$	432	432	432	351	432	351	351	-	-
$N_S$	432	432	432	432	432	432	432	432	432
%>0	100	100	100	81.3	100	81.3	81.3	-	-

For quantity control (Eqs. (24) and (25)), in the first approach, the upper-level problem (decision variable is the toll in tokens;  $\delta_i, i = 1 \dots I$ ) is again solved using GA and implemented using the MATLAB *ga* routine. For each evaluation of the objective function for a given candidate solution of the toll in tokens  $\delta_i, i = 1 \dots I$ , we now need to compute both the equilibrium (SUE) flows ( $X_i^{sk}, i = 0 \dots I, k = 1, 2$ ) and the market clearing prices for each interval ( $p_i^{sk}, i = 1 \dots I, k = 1, 2$ ). Here, we use the iterative price adjustment process described in Section 3.3.4. We start with a candidate set of prices ( $p_i^{sk}, i = 1 \dots I, k = 1, 2$ ), and solve the SUE problem (again using the *lsqnonlin* routine) to determine the flows ( $X_i^{sk}, i = 0 \dots I, k = 1, 2$ ). Next, the demand and supply of tokens ( $S_i^{sk}, D_i^{sk}, i = 1 \dots I, k = 1, 2$ ) are computed and the prices are adjusted using Eq. (18). This process is repeated until convergence of the prices, at which point we have the solutions of ( $X_i^{sk}, p_i^{sk}, i = 1 \dots I, k = 1, 2$ ) for a given toll in tokens ( $\delta_i, i = 1 \dots I$ ), and we can compute the objective function. In this price adjustment process we use a simple bisection method where the value of  $\Delta_p$  in Eq. (16) progressively decreases with successive iterations. In the second approach, once again, the Genetic algorithm is simply replaced by the SQP algorithm.

In all the 432 test instances (for all the instruments), both the GA and the SQP algorithm (with multiple starting points) converged to the same optimal welfare and toll values.

## 5.2. Results and discussion

The results from the numerical experiments and their implications are discussed in this section. After a description of the overall results in terms of optimum social welfare, the effects of congestion, extent of day to day variability, selling model, and heterogeneity are discussed in turn.

### 5.2.1. Welfare

Summary statistics (across the 432 scenarios) of the welfare differences between various instruments under stochastic demand are presented in Table 3. The following abbreviations are used: **NT** for the no-toll equilibrium, **SP** for the price system or tolling in dollars, **SQ** for the quantity system or tolling in tokens. We also include a benchmark (abbreviated **ADP**) in which the tolls (in either dollars or tokens) are adaptive across the two days and set by the regulator based on the realization of demand. Clearly, in this case, the price and quantity instruments are equivalent (as shown in Section 3.4), and this benchmark represents the maximum welfare that can be attained in case of stochastic demand.  $N_S$  denotes the number of scenarios.

The results show that in the case of the fixed token supply, neither instrument is consistently superior across all scenarios (column SQ-SP in Table 3). The quantity system is superior in around 81% of the tested scenarios, with the absolute welfare difference (SQ-SP) ranging between -545\$ and 701\$, and mean and median values of 127 \$ and 90 \$ respectively. To put these differences in context, the total welfare of the no toll equilibrium ranges between 4645\$ and 9021\$ while that of the benchmark ranges between 8024\$ and 10336\$; the total toll revenue in the benchmark system ranges between 4240\$ and 7027\$. The percentage difference in welfare (SQ-SP) relative to the welfare of the no-toll equilibrium ranges between -6.8% and 13.9%, with mean and median percentage differences of 2% and 1.2% respectively. Note that all welfare values (ADP, SQ, SP, NT) can be considered as being relative to a situation where all travelers stay at home (i.e for example, due to very large travel times), which will yield zero welfare due to the normalization of the utility of the cancel trip option. The percentage differences need to be interpreted in this context.

Moreover, an examination of the average welfare difference between the price and quantity instrument (column SQ-SP in Table 3) relative to the average welfare difference between the price instrument and the adaptive benchmark (column ADP-SP in Table 3) suggests that the added flexibility of the credit market allows us to recover a little over 57% of the welfare lost due to the tolls in dollars and tokens being fixed across days. This is also evident when looking at the welfare differences between the price and quantity instruments relative to the adaptive benchmark (column (SQ-SP)/ADP in Table 3), which ranges between -6.2% and 7.2% with a mean value of 1.4%.

The overall distribution of welfare differences (SQ-SP) is shown in Figs. 1(a) and 1(b) (the kernel density is plotted assuming a normal kernel function), which as noted before indicate that when the supply of tokens is fixed across days, neither instrument is consistently superior in terms of efficiency. In order to gain more insights into the conditions under which the quantity instrument is superior, we next examine the impacts of the shape of the congestion function, selling behavior, the benefit difference across days and the extent of heterogeneity.

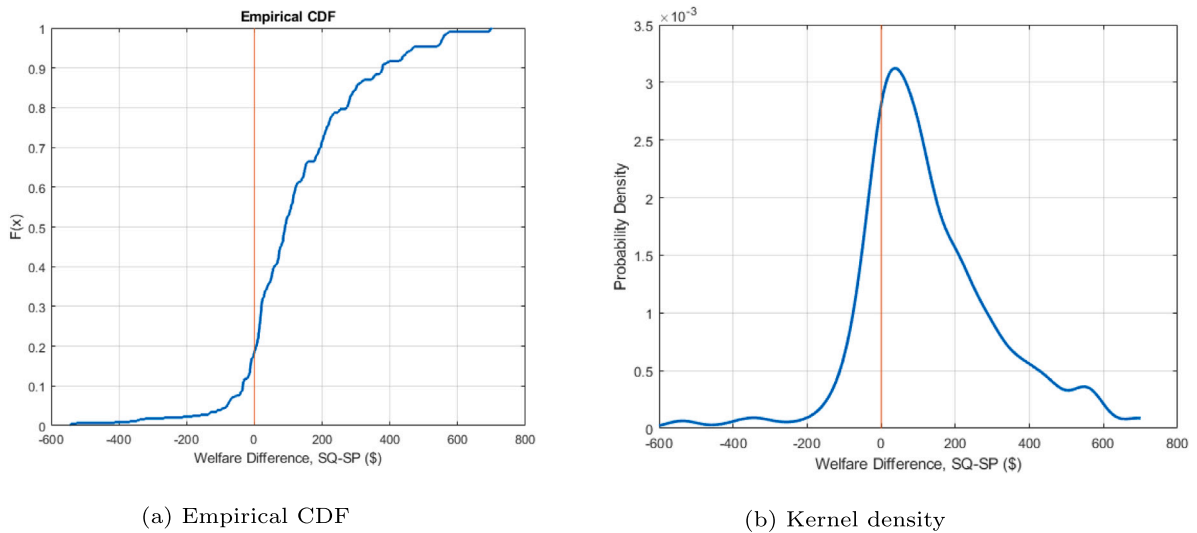


Fig. 1. Distribution of welfare differences.

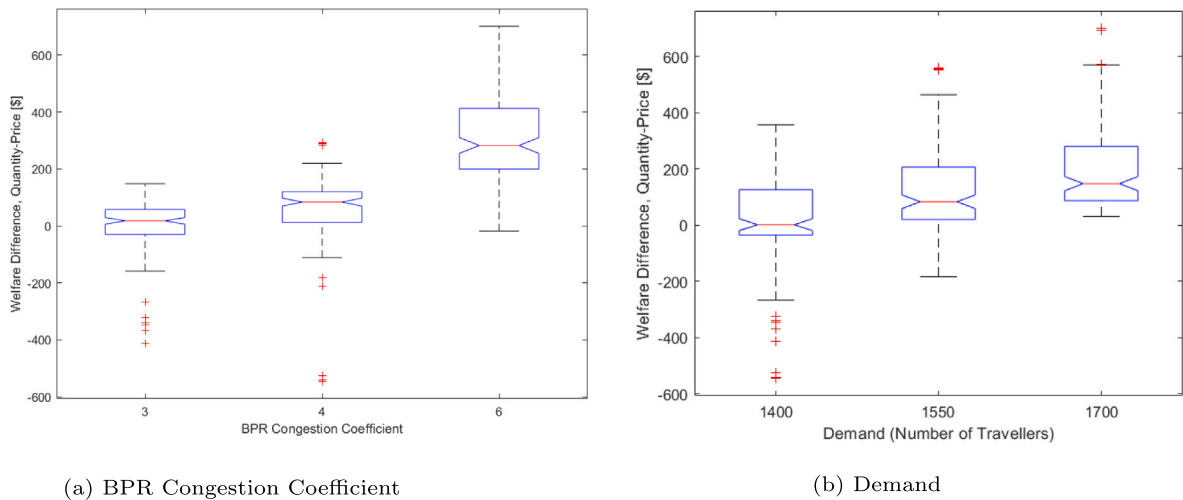


Fig. 2. Welfare difference: Effect of BPR congestion coefficient and demand.

5.2.2. Effect of congestion function

We first examine the effect of the congestion function, and draw on recent theoretical insights from [de Palma and Lindsey \(2020\)](#), who study tradable permit schemes in a setting with homogeneous agents and a single congestible facility. They conjecture (and explore through simple numerical examples) that in the case of variable demand and a fixed, but nonlinear cost function, the performance of a quantity control system dominates that of a price control system when the cost function is more steeply curved. This relates to their general finding that a quantity control system is relatively efficient if optimal usage levels are similar across states whereas a congestion fee achieves high efficiency if the first-best fee varies little over states.

The results from our experiments support these findings. First, the BPR congestion coefficient has a statistically significant effect (level of significance  $\alpha = 0.01$ ) on the difference in total welfare between the quantity and price control systems. The average difference increases from \$2.1 at  $\beta = 3$  to \$63.8 at  $\beta = 4$ , and \$313.9 at  $\beta = 6$ . Moreover, the percentage of scenarios where the quantity control system is superior increases from 61.8% at  $\beta = 3$  to 82.6% at  $\beta = 4$  and 99.3% at  $\beta = 6$ . The effect of  $\beta$  on the welfare difference is shown in the box plot in [Fig. 2\(a\)](#) where it can be seen that the quantity control mechanism is superior in terms of total welfare typically when the congestion curve is steeper or more convex (high value of  $\beta = 4, 6$ ). In the box plot, the lower and upper edges of the blue box represent the 25th and 75th percentile respectively, the red line represents the median, and the notch represents a 95% confidence interval for the median.

A similar trend of increasing and statistically significant ( $\alpha = 0.01$ ) welfare differences (SQ-SP) is observed as the demand level increases (under fixed capacity, i.e., congestion levels increase). The average difference increases from \$26.9 at  $N = 1400$  to \$138.6



**Table 4**  
Illustrative scenarios: Effect of BPR congestion coefficient  $\beta$ .

Instrument	Day	Flows				Optimal/Equivalent tolls		
		$T_1$	$T_2$	$T_3$	$T_0$	$T_1$	$T_2$	$T_3$
<b>Scenario 1</b>								
No Toll (NT)	s1	377	542	367	115	0	0	0
	s2	419	567	407	8			
Stochastic Price (SP)	s1	326	404	315	356	2.59	4.31	2.63
	s2	436	498	421	45			
Stochastic Quantity (SQ)	s1	374	459	351	216	1.16	2.67	1.33
	s2	395	428	365	212			
Benchmark (ADP)	s1	338	416	327	319	2.26	3.96	2.30
	s2	411	474	397	118			
<b>Scenario 2</b>								
No Toll (NT)	s1	389	469	382	160	0	0	0
	s2	441	498	432	28			
Stochastic Price (SP)	s1	287	337	281	495	3.17	4.90	3.21
	s2	419	443	410	128			
Stochastic Quantity (SQ)	s1	343	401	340	316	1.52	2.84	1.63
	s2	360	372	355	313			
Benchmark (ADP)	s1	326	366	319	389	2.59	4.30	2.63
	s2	372	401	364	263			

at  $N = 1550$ , and 214.4\$ at  $N = 1700$ . Moreover, the percentage of scenarios where the quantity control system is superior increases from 50.7% at  $N = 1400$  to 93.1% at  $N = 1550$  and 100% at  $N = 1700$ . The effect of total demand (number of travelers) on the welfare differences between the quantity and price instruments is shown in the box plot in Fig. 2(b).

In order to gain more insight into the effect of the BPR congestion coefficient and the total demand (number of users), we examine several illustrative scenarios. First, we compare two scenarios (referred to as 1 and 2) with  $\beta = 3, N = 1400, \Delta = 5$  and  $\beta = 6, N = 1400, \Delta = 5$ , respectively. All other factors including the scale of the mobility model and selling model are the same. The welfare difference between the quantity and price instruments (SQ-SP) are  $-72.9\$$  and  $271.0\$$  in scenario 1 and 2 respectively.

Table 4 summarizes the flows in different time periods and the tolls in dollar amounts (note that for the price instrument this is directly the toll in dollars whereas for the quantity instrument it is the product of the toll in tokens and the token market price). First, observe that for both scenarios, as expected, under the price instrument (SP), the number of individuals traveling (total flow in periods  $T_1, T_2, T_3$ ) varies significantly across the days s1 and s2 (also evident from the number of travelers canceling trip, i.e., flow in  $T_0$ ) whereas the toll in dollar amounts is fixed. In contrast, under the quantity instrument, the number of travelers traveling is roughly the same across the two days whereas the toll in dollar amounts varies significantly. Next, we see that the optimal usage of the network (or number of people traveling) under the adaptive benchmark ADP varies more across the states S1 and S2 in scenario 1 than in scenario 2 (difference in the optimal flows for  $T_0$  across days is  $319 - 118 = 201$  for scenario 1 versus  $389 - 263 = 126$  in scenario 2). Conversely, looking at optimal tolls under the benchmark ADP, one can see that the toll difference across s1 and s2 is higher in scenario 2 compared to scenario 1 (in interval  $T_2$ ,  $7.06 - 4.30 = 2.76$  for scenario 2 versus  $6.36 - 3.96 = 2.40$  for scenario 1). Thus, the results suggest that at higher BPR coefficients or steeper congestion functions, the optimal usage levels of the network are relatively more similar across states leading to superiority of the quantity instrument. More intuition for this is provided analytically by de Palma and Lindsey (2020) who look at the welfare losses of the two instruments using a single congested alternative under linear and non-linear demand.

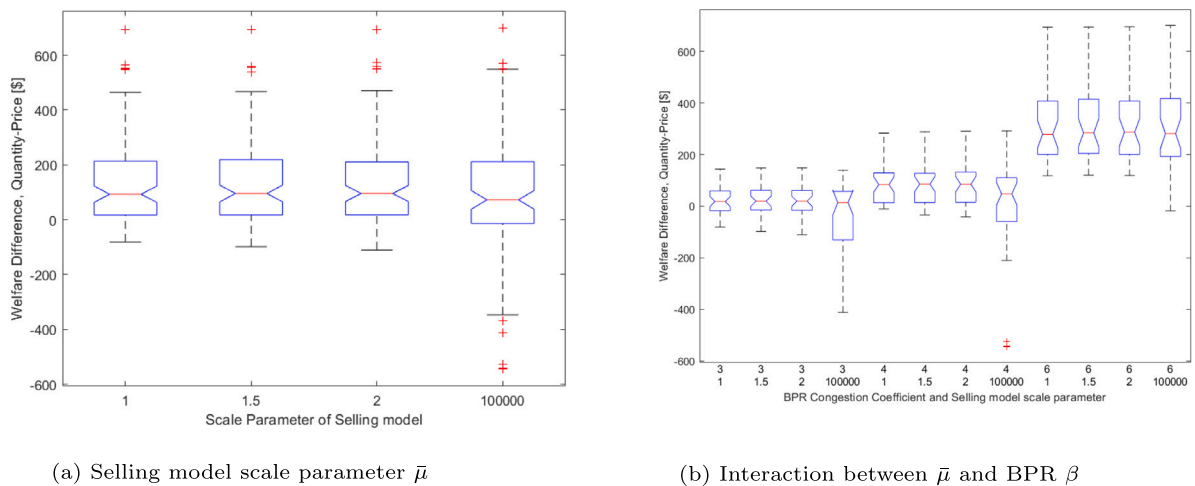
Along similar lines, to examine the effect of the total number of travelers  $N$ , we compare two different scenarios (referred to as 1 and 2) with  $\beta = 3, N = 1400, \Delta = 5$  and  $\beta = 3, N = 1700, \Delta = 5$  respectively. All other factors including the scale of the mobility model and selling model are the same. The welfare difference between the quantity and price instruments (SQ-SP) are  $-72.9\$$  and  $88.73\$$  in scenario 1 and 2, respectively. Table 5 presents the flows in different time periods and the tolls in dollar amounts for these two scenarios. We observe – as before in the case of higher  $\beta$  – that when the overall demand level is higher (scenario 2), the difference in optimal toll rates across s1 and s2 is higher compared to scenario 1 (in interval  $T_2$ ,  $7.55 - 4.57 = 2.98$  for scenario 2 versus  $6.36 - 3.96 = 2.40$  for scenario 1). Thus, at higher demand levels (and hence, more severe congestion effects) the optimal toll rates vary more across states, once again leading to superior performance of the quantity control instrument relative to scenarios with lower demand levels.

### 5.2.3. Selling model

The explicit treatment of selling behavior is an important characteristic of the proposed model and allows us to examine the impact of the selling decisions on the performance of the quantity control system. Fig. 3(a) presents a box plot of the effect of the scale parameter of the selling model on the difference between the quantity and price instruments. The value of  $\bar{\mu} = 100,000$ , which corresponds to  $\bar{\mu} \rightarrow \infty$  represents a purely random selling model (or a non-rational market) and has the impact of equalizing the supply of tokens across the three time periods (consequently, the demand of tokens as well). As the results show, this has the effect

**Table 5**  
Illustrative scenarios: Effect of demand.

Instrument	Day	Flows				Optimal tolls		
		$T_1$	$T_2$	$T_3$	$T_0$	$T_1$	$T_2$	$T_3$
<b>Scenario 1</b>								
No Toll (NT)	s1	377	542	367	115	0	0	0
	s2	419	567	407	8			
Stochastic Price (SP)	s1	326	404	315	356	2.59	4.31	2.63
	s2	436	498	421	45			
Stochastic Quantity (SQ)	s1	374	459	351	216	1.16	2.67	1.33
	s2	395	428	365	212			
Benchmark (ADP)	s1	338	416	327	319	2.26	3.96	2.30
	s2	411	474	397	118			
<b>Scenario 2</b>								
No Toll (NT)	s1	446	586	432	235	0	0	0
	s2	522	638	505	35			
Stochastic Price (SP)	s1	321	399	311	668	3.65	5.39	3.69
	s2	493	549	476	182			
Stochastic Quantity (SQ)	s1	399	498	387	416	1.85	3.04	1.86
	s2	416	469	401	414			
Benchmark (ADP)	s1	362	435	350	553	2.84	4.57	2.88
	s2	441	500	426	333			



**Fig. 3.** Welfare difference: Effect of selling model.

of a deterioration in the performance of the quantity control system, reflected in the mean difference in welfare between the two instruments (Quantity – Price), which takes a mean value of 85.7 S\$ at  $\bar{\mu} = 100\,000$  versus 140.2 S\$ for  $\bar{\mu} \in [1, 2]$ . Note that within the range of  $\bar{\mu} \in [1, 2]$ , performance of the quantity control does not vary substantially. However, interestingly, even within the case of a non-rational market, the quantity instrument remains superior to the price instrument in cases where congestion effects are severe (e.g.  $\beta = 6$  in Fig. 3(b)). In other words, the advantages of the quantity control system noted in Section 5.2.2 remain even if users are not perfectly rational in the selling market, although the extent of welfare difference is marginally lower.

The impact of market behavior is illustrated in Fig. 4 for a few selected scenarios which yield the highest deterioration in welfare for  $\bar{\mu} = 100\,000$  compared to the corresponding scenarios with  $\bar{\mu} = 1, 1.5, 2$ . Each marker type or series represents scenarios where all parameters are identical except  $\bar{\mu}$ . Note that the interpolating lines between each point are not meant to be indicative of the actual trend but are used to simply make the figure more legible. As noted above, we observe that within the range  $\bar{\mu} \in [1, 2]$ , the differences in welfare of the quantity instrument are negligible whereas the welfare deteriorates if users are more irrational in the selling market. Thus, the findings suggest that market design aspects of the quantity control instrument are important and can have effects on efficiency. This is explained in more detail next.

In order to gain more intuition into the effect of selling behavior in the market, we examine two illustrative scenarios, one with  $\bar{\mu} = 1$  and the second with  $\bar{\mu} = 100\,000$  ( $\beta = 4, N = 1550, \Delta = 3$  in both scenarios). Thus, the two scenarios are identical in all respects except the scale parameter of the selling model. The welfare of the scenario with  $\bar{\mu} = 1$  is higher than that with  $\bar{\mu} = 100\,000$  by

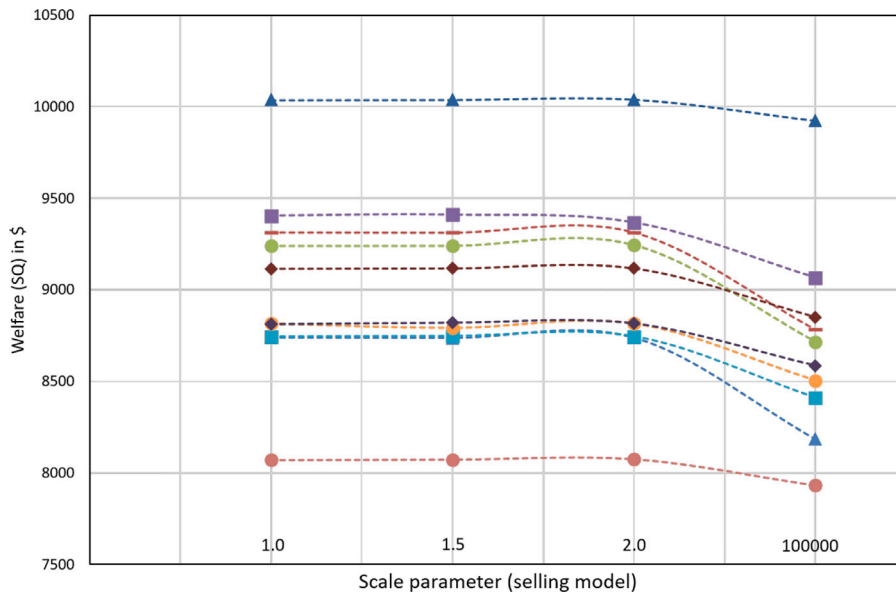


Fig. 4. Effect of  $\bar{\mu}$  on welfare of quantity control (SQ) for selected scenarios (note: each series represents a given scenario with varying  $\bar{\mu}$ ).

Table 6

Illustrative scenarios: Effect of selling behavior.

Instrument	Day	Flows				Optimal tolls		
		T1	T2	T3	T0	T1	T2	T3
Stochastic Quantity (SQ) $\bar{\mu} = 1.0$	s1	366	440	353	391	2.22	3.77	2.29
	s2	381	418	365	386	5.00	7.10	5.09
Stochastic Quantity (SQ) $\bar{\mu} = 100000$	s1	355	378	341	476	2.72	5.06	2.82
	s2	355	378	341	476	5.72	8.06	5.82
Benchmark (ADP)	s1	346	409	336	459	2.26	3.96	2.30
	s2	391	444	380	335	4.60	6.36	4.64
		Demand for tokens				Supply of tokens		
		T1	T2	T3	T0	T1	T2	T3
Stochastic Quantity (SQ) $\bar{\mu} = 1.0$	s1	53.89	148.93	56.04	-	53.89	148.93	56.04
	s2	56.07	141.53	57.96	-	56.05	141.53	57.95
Stochastic Quantity (SQ) $\bar{\mu} = 100000$	s1	111.81	111.81	111.81	-	111.81	111.81	111.81
	s2	111.81	111.81	111.81	-	111.81	111.81	111.81

104.2\$. Table 6 summarizes the flows in different time periods, the tolls in dollar amounts (note that is the product of the toll in tokens and the token market price) and the demand and supply of tokens for all time intervals and both days s1 and s2. First, note that non-rational selling behavior or  $\bar{\mu} = 100000$  has the effect of equalizing the probability of selling in all three time intervals and hence, equalizes the supply of tokens for all three intervals (last two rows in Table 6). Interestingly, this also causes the token supply to be equal on both days s1 and s2 (for all three intervals). Thus, we see that the optimal tolls (SQ with  $\bar{\mu} = 100000$ ) on days s1 and s2 differ by an additive constant of 3\$ (which is exactly equal to the benefit difference between the two days,  $\Delta = 3$ ) resulting in identical token supply and also, identical flows on both days s1 and s2. Moreover, the equal token supply results in lower flows in interval  $T_2$  (for example, 440 on s1 for  $\bar{\mu} = 1$  versus 378 on s1 for  $\bar{\mu} = 100000$ ). This results in a significantly higher number of travelers canceling trip (choosing  $T_0$ ) in the scenario with  $\bar{\mu} = 100000$  (476) versus the scenario with  $\bar{\mu} = 1$  (391 and 386) leading to a loss in welfare. In summary, we see that the quantity control instrument is robust with respect to selling behavior in the market and is still superior even with irrational sellers when congestion effects are more severe. However, there is a deterioration in welfare when the behavior of sellers is more irrational, which causes the equilibrium price of tokens to increase leading to less travel.

#### 5.2.4. Difference between states of nature

The benefit difference between the two days or states of nature ( $\Delta$ ) is a measure of the extent of day to day variability, and the results indicate – similar to the congestion coefficient and demand – that it significantly affects the relative performance of the two instruments ( $\alpha = 0.05$ ). The average difference increases from \$81 at  $\Delta = 3$  to \$125 at  $\Delta = 4$ , and 175.4\$ at  $\Delta = 5$ .

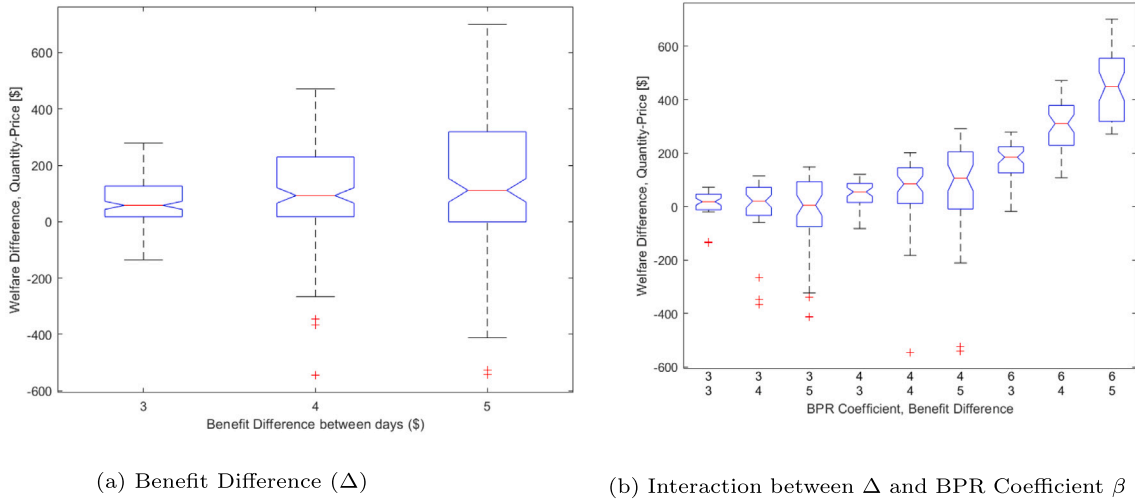


Fig. 5. Welfare difference: Effect of benefit difference and interaction between benefit difference/beta.

Further, the variance in difference between the two instruments also increases as is evident from the boxplot in Fig. 5(a). This can be better understood by examining the interaction effects of the benefit difference with the BPR congestion coefficient. Thus, when the congestion curve is steeper, a larger degree of day to day variability results in a greater advantage for the quantity control system ( $\beta = 6, \Delta = 5$  versus  $\beta = 6, \Delta = 3$  in Fig. 5(b)). On the hand, when congestion effects are less severe, an increase in day to day variation ( $\beta = 3, \Delta = 5$  versus  $\beta = 3, \Delta = 3$  in Fig. 5(b)) results in a poorer performance of the quantity control system.

This can be explained once again by the fact that at higher levels of the benefit difference (or day to day variability) and BPR coefficient, the optimal toll rates (in dollar amounts) tend to vary significantly across the states  $s_1$  and  $s_2$ , leading to superior performance of the quantity instrument.

### 5.2.5. Impacts of heterogeneity

The impact of the coefficient of variation (COV) of the scale parameter  $\mu_n$  in the mobility model (note that  $\mu_n$  is assumed to be lognormally distributed across the population of travelers) is shown in the boxplot in Fig. 6. First, it can be observed that as the COV increases from  $\mu_n = 0.2$  to  $\mu_n = 0.5$ , the mean difference in welfare between the quantity and price instruments increases only marginally from 109.8 \$ to 123.1 \$ (statistically insignificant at  $\alpha = 0.05$ ).

In contrast, when we examine the effect of overall heterogeneity in the mobility model (recall that the scenarios with  $\mu_n = 0$  represent the homogeneous case where all other parameters in the mobility model are also assumed to be deterministic), we see a significant effect ( $\alpha = 0.05$ ). The mean difference between the quantity and price control systems for the homogeneous scenarios is in fact higher at 158\$ (scenarios with COV of  $\mu_n = 0$  in Fig. 6) compared to 115\$ when heterogeneity is considered (scenarios with COV of  $\mu_n = 0.2, 0.33, 0.5$  in Fig. 6). This has important implications and suggests that ignoring heterogeneity can potentially overestimate the benefits of the quantity control system. This contrasts with the findings in de Palma et al. (2018) where heterogeneity was found to slightly increase the average welfare difference between the two instruments. The intuition for these differences is hard to arrive at; one potential cause may be differences in the nature of variability, which arises from scheduling preferences in our case as opposed to the total number of users in their case.

### 5.2.6. Equity: winners and losers

Equity is a key consideration in the comparison between the price and quantity control instruments. In this section, we compare the two instruments using the Gini coefficient computed based on the logsum (a measure of user benefits). For a detailed discussion of measures of inequality and welfare in the transportation context, we refer the readers to Trannoy (2011) and Delle Site et al. (2021). We first discuss the computation of the Gini coefficient followed by a discussion of findings.

Consider the population of  $N$  travelers ( $k = 1 \dots N$ ), and let  $UB(k)$  denote the user benefit of individual  $k$  in \$ amounts. In the case of the no toll equilibrium (denoted NT) and tolling in dollars with no redistribution of toll revenues (denoted SPN),  $UB(k)$  is simply the logsum of individual  $k$ . In the case of tolling in dollars with an equal redistribution of toll revenues (denoted SP),  $UB(k)$  is the logsum of individual  $k$  plus the average toll revenue per individual. Finally, in the case of tolling in tokens (denoted SQ),  $UB(k)$  is the logsum of individual  $k$  plus the market value of the initial token endowment (since tokens are distributed for free and no tokens are unused at the end of the day). Assume that drivers arranged in increasing order of their user benefit and let  $x = k/N$ . Define,

$$g(x) = \frac{\sum_{j=1}^{xN} UB(j)}{\sum_{j=1}^N UB(j)}, \tag{26}$$

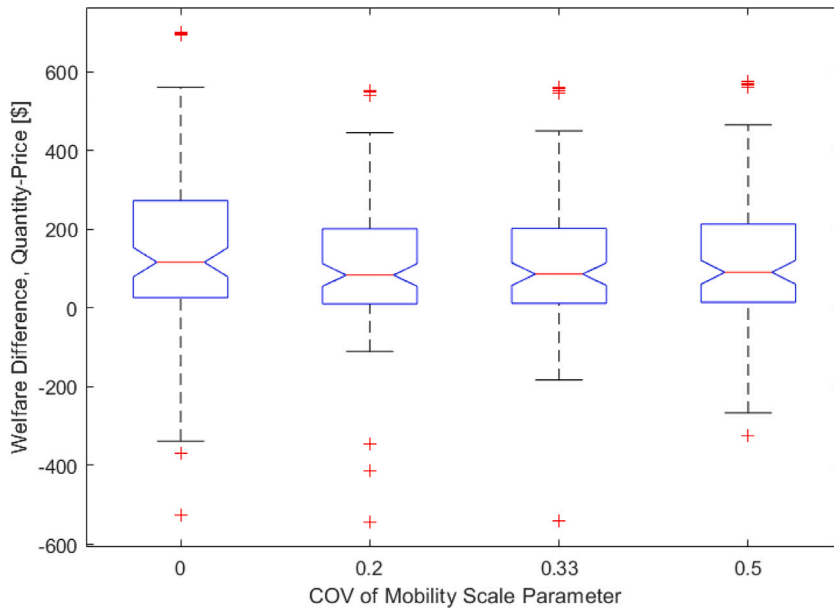


Fig. 6. Welfare difference: Effect of heterogeneity and COV of mobility scale.

Table 7  
Summary statistics: Gini coefficient.

Statistic	$GC_{NT}$	Percentage difference in Gini coefficient			
		(SPN-NT)/NT	(SP-NT)/NT	(SQ-NT)/NT	(SQ-SP)/SP
Mean	0.183	28.7	-43.7	-43.7	-0.6
Median	0.175	32.3	-43.0	-43.9	0.0
Minimum	0.123	-2.4	-65.9	-68.9	-11.2
Maximum	0.275	46.2	-22.9	-1.9	40.1
25th percentile	0.155	21.7	-51.8	-52.2	-3.3
75th percentile	0.208	38.1	-34.2	-34.3	1.4
Scenarios >0	324	312	0	0	156
% Scenarios >0	100	96.3	0	0	48.1
$N_S$	324				

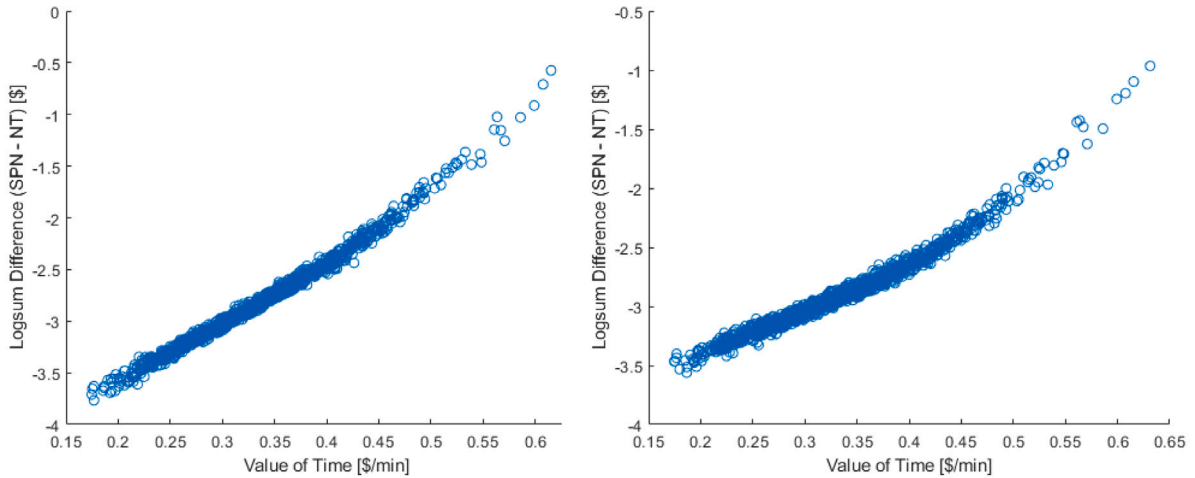
where  $g(x)$  represents the Lorenz curve, which is the cumulative share of total user benefits (based on the logsum measure) obtained by the bottom  $xN$  individuals in the population (note that  $xN$  is an integer). The Gini coefficient of user benefits (denoted GC) is computed as,

$$GC = \frac{\left| 0.5 - \int_0^1 g(x)dx \right|}{0.5} \tag{27}$$

The Gini coefficient is a measure of equity and takes a value between 0 and 1; a value equal to zero implies total equity and a value of 1 indicates total inequity. The larger it is, the more inequitable is the policy.

Table 7 summarizes the distribution of the Gini coefficient (across the 324 scenarios with heterogeneity) for the NT equilibrium (denoted by  $GC_{NT}$ ) and percentage differences between the Gini coefficient for the different instruments. First, observe that with tolling in dollars wherein toll revenues are not redistributed, in a majority of the scenarios (96.3%), the Gini coefficient increases (i.e., is more inequitable) relative to the No Toll equilibrium and is on average 28.7% higher (column three of Table 7). This is in line with the general observation that pricing is vertically inequitable and benefits the rich (here the individuals with high value of time) more than the poor. The scatter plots (Fig. 7) of logsum difference (between SP and NT) versus value of time (a proxy for income) corroborate this observation, where we see that the benefits clearly increase with an increase in value of time. The plots represent two illustrative scenarios for the s2 day and each point in the plot represents an individual. Interestingly, there are a small number of scenarios (3.7%), where the Gini coefficient reduces even when toll revenues are not redistributed. This occurs in scenarios where the congestion effects are the most severe (BPR coefficient of 6 and highest demand level).

Our second observation is that both in case of the quantity instrument and the price instrument (when toll revenues are equally redistributed), there is a significant improvement of equity, by an average of 43.7%, relative to the No Toll equilibrium (column four and five of Table 7). This is a key finding and implies that both instruments favor the poor (low value of time individuals)



(a) Scenario 1 ( $\beta = 6, N = 1400, \Delta = 5$ )

(b) Scenario 2 ( $\beta = 3, N = 1550, \Delta = 4$ )

Fig. 7. User benefits versus value of time.

Table 8

Summary statistics: Welfare differences (Adaptive token supply).

Statistic	Welfare difference (\$)					Percentage diff. (SQ_A-SP)/NT	Welfare (\$) NT
	SP-NT	SQ_A-NT	SQ-SP	SQ_A-SP	ADP-SP		
Mean	1555	1745	127	190	222	2.9	7347
Median	1416	1561	90	158	169	2.1	7443
Min	517	348	-545	-177	39	-2.0	4645
Max	3890	4656	701	767	778	15.2	9021
25th per.	867	952	15	82	103	1.0	6613
75th per.	1896	2143	213	263	286	4.0	8131
$N_s > 0$	432	432	351	410	432	410	-
%>0	100	100	81.3	94.9	100	94.9	-

since there is a large reduction in the Gini coefficient across all scenarios relative to the No Toll equilibrium. The primary reason is that an equal redistribution of toll revenues (price instrument) and the equal allocation of tokens (quantity instrument) results in an increase in the cumulative share of benefits obtained by lower income travelers, leading to an improvement in equity. Further, note that in case of the quantity instrument, a further improvement in equity can be achieved through any progressive allocation of the tokens.

Finally, comparing the Gini coefficient for the quantity instrument and the price instrument with equal redistribution of toll revenues (column six of Table 7), we see that neither instrument is consistently superior in terms of equity, although the quantity instrument is on average marginally better (average difference of 0.6%). Moreover, similar to the comparative performance with respect to welfare, we find that the quantity instrument is superior in terms of equity in scenarios with more severe congestion effects (high BPR congestion coefficient of 4 and 6) and when the selling behavior of individuals is rational. These findings make an additional case for tolling in tokens.

### 5.3. Adaptive token supply

The experiments in Section 5.2 consider a quantity control system wherein the token supply is fixed across days and the results indicate the tolling in tokens is not consistently superior to tolling in dollars. The results also suggest that the price system is typically superior when congestion effects are less severe (slope of the congestion function is less steep, demand is lower). In these cases, as seen in the illustrative scenarios in Section 5.2.2, the quantity targets may be too lax on the s1 day and thus, the performance of the quantity control system can be improved by allowing the token supply to be adapted across days in response to the realization of demand (Eq. (25)).

In this section, we examine the comparative performance of the two instruments when the token supply is adaptive. Note that, in this case certain parameters of the quantity control system ( $M^{s1}, M^{s2}$  in Eq. (25)) are dependent on the state of nature. This is in contrast with the adaptive tradable permit system (TPS) considered in de Palma and Lindsey (2020) where the regulator issues a certain number of permits, but in addition, offers to sell further permits at a price  $s$ , and buy permits at a price  $r$ , where  $r < s$ . This limits the price of permits to the range  $[r, s]$ , where  $r$  and  $s$  are fixed and state independent.



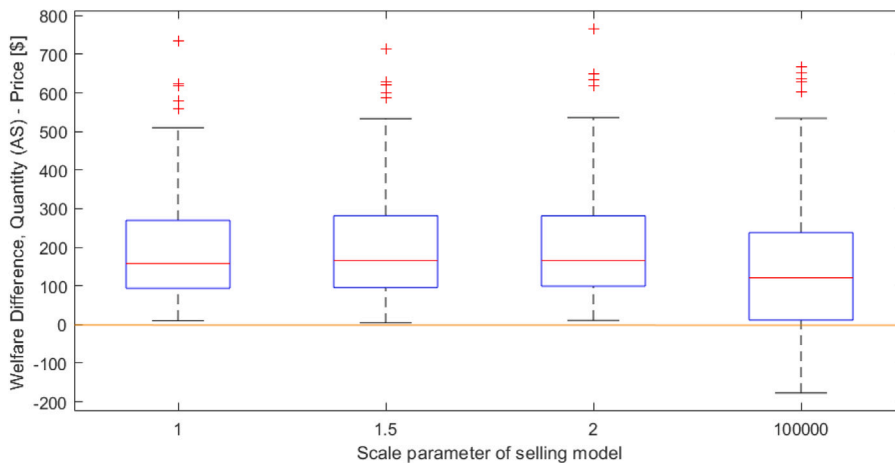


Fig. 8. Welfare differences (adaptive token supply): Effect of selling model.

The results are summarized in Table 8 and as expected, indicate that the quantity control system with adaptive supply is superior to that with fixed supply in all scenarios. The mean welfare difference between the quantity system with adaptive supply (denoted SQ\_A) and the price system is 190 \$ compared to 127 \$ with fixed supply (refer columns SQ-SP and SQ\_A-SP). Moreover, a comparison of these numbers against the mean difference of 222\$ between the adaptive benchmark and the price system (ADP-SP) reveals the extent of welfare improvements that can be attained by adapting the token supply across days. Thus, while the quantity instrument with fixed token supply recovers a little over 57% of the welfare loss due to fixing the tolls (in dollars and tokens) across days, the quantity instrument with adaptive token supply recovers almost 86% of this welfare loss.

However, contrary to intuition, even with adaptive token supply, the quantity control system is still not consistently superior to the price control system although it yields a higher welfare in 94.9% of the tested scenarios. This is in contrast with the findings in de Palma et al. (2018) for a single period setting where the quantity control with adaptive token supply is consistently superior. A more detailed examination shows that the scenarios where the price control is superior are in fact all scenarios where the selling behavior is non-rational or completely random (i.e.,  $\bar{\mu} \rightarrow \infty$ ), as shown in the box plot in Fig. 8. As shown in Section 5.2.3, this has the effect of forcing the token supply to be equal across time periods, reducing the efficiency of the quantity control system in a manner that is not redressed even with the adaptive token supply. This once again highlights the importance of market design in the efficiency of the quantity control system.

## 6. Conclusions and further research

This paper develops a methodology to compare price control (tolling in dollars) and quantity control (tolling in tokens) instruments in the context of a within-day setting with departure time choice. In the quantity control system, users are provided an initial endowment of tokens by the regulator and incur a token charge to travel in a specific time period. Tokens can be bought and sold within a marketplace at a price determined by a smooth market clearing mechanism in each time period. A key feature of the market model is that the dynamics of price adjustment and selling decisions of users are explicitly considered.

Numerical experiments across a wide range of scenarios with demand uncertainty yield the following key insights. First, when the tolls (in dollars and tokens) can be adapted from day to day, the two instruments are equivalent. Second, when the token supply is fixed across days or states of nature and the tolls (in dollars and tokens) are non-adaptive, the quantity control instrument is superior in welfare terms when congestion effects are more severe, i.e., steep congestion functions (realistic BPR models), high demand levels and high day-to-day variability. In these scenarios, the optimal network usage is relatively similar across states whereas the optimal toll in dollar amounts varies significantly across states. Third, non-rational selling behavior, which has the effect of equalizing token supply across time intervals leads to a deterioration in the performance of the quantity instrument. However, in general, the token system is robust (in welfare terms) with respect to selling behavior in the market. Fourth, when the token supply can be adapted from day-to-day, the quantity instrument is superior in all scenarios where selling behavior is rational. Finally, when toll revenues in the price instrument are equally redistributed (typically difficult to implement in practice) and tokens (in the quantity instrument) are equally distributed, tolling in tokens is marginally more equitable in scenarios where congestion effects are more severe. These findings make a potential case for quantity control.

Several points are however noteworthy. First, income effects and second-order effects on the use of toll revenues are not considered. Second, transaction costs associated with the trading of credits, the process of finding a buyer or seller, negotiating a price, etc. are ignored. These are likely to affect the overall welfare of the quantity control system (see Nie (2012)). However, as noted by Brands et al. (2020), transaction costs may be minimized through suitable market designs. For instance, they make use of a price setting intermediary with whom users trade, and point out that this can significantly reduce transaction and negotiation

costs compared to designs that include consumer to consumer trading (and over existing designs such as Dutch and English auctions, sealed-bid auctions and Vickrey auction markets). Third, the public acceptability of tradable credits is not necessarily guaranteed and will depend on the initial allocation of credits and the extent of volatility in the credit market.

There are several avenues of further research including the use of more realistic network and congestion models (for example, the morning commute problem with  $\alpha - \beta - \gamma$  preferences (Lamotte and Geroliminis, 2021) and general networks), the consideration of both departure time and route/mode choice, joint modeling of the travel and selling decision, and the inclusion of income effects and transaction costs. Further, the model can be extended to allow for users to buy tokens and resell them later if they change their travel decision, a market feature that is clearly desirable.

### CRedit authorship contribution statement

**Ravi Seshadri:** Conceptualization, Methodology, Visualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **André de Palma:** Conceptualization, Methodology, Investigation, Formal analysis, Supervision, Funding, Writing – original draft, Writing – review & editing. **Moshe Ben-Akiva:** Conceptualization, Methodology, Supervision, Funding, Writing – review & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

André de Palma was partially supported by the NSFC-JPI UE joint research project “MAAT” (project no. 18356856), Affinité, ANR, France project and by Labex MME-DII (France). Part of this research was funded by the U.S. National Science Foundation (Award CMMI-1917891). We would like to gratefully acknowledge the beneficial comments and suggestions of Robin Lindsey. We also acknowledge the illuminating comments and constructive suggestions of the editor and two anonymous referees who helped greatly improve the content and clarity of the manuscript.

### Appendix A. Existence, uniqueness and convergence of price adjustment schemes

Recall that the market and network equilibrium conditions for the quantity control system (deterministic case in Section 3) can be written as,

$$\begin{cases} X_i = \sum_{n=1}^N \frac{\exp[V^n(T_i, \theta_n)]}{\sum_j \exp[V^n(T_j, \theta_n)]}, i = 1 \dots I, \\ X_i (\delta_i - M) = Q_i \left( \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M \right), \forall i \in \tilde{I}, \end{cases} \quad (28)$$

where  $X_i \in D$ ,  $D = (X : \sum_{i=0 \dots I} X_i = N; X_i \geq 0, i = 0 \dots I)$ ;  $Q_i$  is given by Eq. (9). Define  $\tilde{p} = (p_i, i \in \tilde{I})$  and  $\mathbf{X} = (X_0, X_1, \dots, X_I)$ . We wish to show that there exists a solution  $(X, \tilde{p})$  to the above equilibrium conditions (note that  $p_i = 0, \forall i \notin \tilde{I}$ ). We denote the demand and supply of tokens in interval  $i \in \tilde{I}$ , as  $D_i$  and  $S_i$  respectively. Thus,  $D_i = X_i (\delta_i - M)$ , and  $S_i = Q_i \left( \sum_{j \in I \setminus \tilde{I}} X_j (M - \delta_j) + X_0 M \right)$ . Note that  $\tilde{p} > 0$ , since we have assumed that  $\sum_{i=1 \dots I} \tilde{X}_i \delta_i > MN$ , where  $\tilde{X}_i$  ( $i = 0 \dots I$ ) denotes the equilibrium flows in the absence of tolls (see Remark 3.1 in Ye and Yang (2013)).

From the definition of  $h(\cdot)$  in the price adjustment mechanism in Eq. (17), at equilibrium, we have  $h(D_i - S_i) = 0, \forall i \in \tilde{I} \iff D_i = S_i, \forall i \in \tilde{I}$ . From Lemma A.1 in Ye and Yang (2013),  $(X, \tilde{p})$  solves  $h(D_i - S_i) = 0, \forall i \in \tilde{I}$  if and only if it solves  $p_i = [p_i + \rho \cdot (D_i - S_i)]^+, \forall i \in \tilde{I}$ , where  $[x]^+ = \max[x, 0]$  and  $\rho > 0$ . Thus, the market and network equilibrium conditions in Eq. (28) are equivalent to the following fixed-point problem:

$$\begin{cases} X_i = \sum_{n=1}^N \frac{\exp[V^n(T_i, \theta_n)]}{\sum_j \exp[V^n(T_j, \theta_n)]}, i = 1 \dots I, \\ p_i = [p_i + \rho \cdot (D_i - S_i)]^+, \forall i \in \tilde{I}. \end{cases} \quad (29)$$

Along the lines of Ye and Yang (2013), we can demonstrate that the fixed-point problem in Eq. (29) has a solution by invoking the Brouwer–Kakutani fixed-point theorem.

First, note that  $[\cdot]^+$  is continuous. Since  $t_i(X_i)$  are continuous functions of  $X_i$  for  $i = 1 \dots I$ , the right hand side of the first set of equations in (29) are continuous functions of  $(X, \tilde{p})$ . Further observe that the set of demand feasible flows  $D$  forms a closed and convex set. Thus, it suffices to show that there exists closed and convex sets  $\Omega_{p_i}, \forall i \in \tilde{I}$  such that for all  $p_i \in \Omega_{p_i}, [p_i + \rho \cdot (D_i - S_i)]^+ \in \Omega_{p_i}$ . Observe also that since we have a toll free alternative  $T_0$ ,

$$\lim_{p \rightarrow \infty} \sum_{i \in \tilde{I}} D_i \rightarrow 0, \quad (30)$$

we do not require a lower bound on the token allocation to guarantee solution existence (as is necessary in [Ye and Yang \(2013\)](#)).

Denote  $Z_i(X, \bar{p}) = D_i - S_i, i \in \bar{I}$ . For some  $\hat{X} \in D$  (denote this set as  $D^-$ ), if  $Z_i(\hat{X}, \bar{p}) \leq 0, \forall \bar{p} > 0$ , then  $[p_i + \rho \cdot (D_i - S_i)]^+ \leq p_i, \forall \bar{p} > 0$ . Next consider  $\hat{X} \in D \setminus D^-$  for which  $Z_i(\hat{X}, \bar{p}) > 0$  for some  $\bar{p} > 0$ . For convenience, we expand  $\bar{p}$  as  $\bar{p} = (p_i, \bar{p}^*)$ , where  $\bar{p}^* = (p_j, j \neq i, j \in \bar{I})$ . Clearly, from Eq. (30) (and given that  $D_i - S_i$  is strictly decreasing in  $p_i$ ), there exists some  $\bar{p}_i > 0$ , such that  $Z_i(\hat{X}, (p_i, \bar{p}^*)) \leq 0, \forall p_i \geq \bar{p}_i$ , and thus,  $[p_i + \rho \cdot Z_i(\hat{X}, (p_i, \bar{p}^*))]^+ \leq p_i, \forall p_i \geq \bar{p}_i$ . Thus, if we define,  $p_i^+ = \max_{p_i \leq \bar{p}_i} [p_i + \rho \cdot Z_i(\hat{X}, (p_i, \bar{p}^*))]^+$ , then  $\forall p_i \in [0, p_i^+], [p_i + \rho \cdot Z_i(\hat{X}, (p_i, \bar{p}^*))]^+ \leq p_i^+$ . Hence,  $\forall p_i \in \Omega_{p_i} \triangleq [0, \max_{\hat{X} \in D \setminus D^-, \bar{p}^* > 0} p_i^+], \forall i \in \bar{I}$ , and  $X \in D, [p_i + \rho \cdot Z_i(X, \bar{p})]^+ \in \Omega_{p_i}, \forall i \in \bar{I}$ .

Thus, from the above, for intervals  $i_1, i_2, \dots, i_{|\bar{I}|} \in \bar{I}, D \times \Omega_{p_{i_1}} \times \dots \times \Omega_{p_{i_{|\bar{I}|}}}$  is a compact and convex set, implying that there exists a solution to the fixed-point problem in Eq. (29). This completes the proof.

Demonstrating solution uniqueness analytically is less straightforward and hence, we examine it numerically using six arbitrarily chosen example test instances (COV of  $\mu_n = 0.2, \bar{\mu} = 1.5, N = \{1550, 1700\}, \beta = \{3, 4, 6\}$ ).  $\delta_i, i = 1 \dots I$  and  $M$  are chosen arbitrarily. For simplicity, we use the deterministic model in Section 3. We first consider the price adjustment mechanism in Eq. (17), where  $h(D_i - S_i)$  is defined as  $h(D_i - S_i) = 0.01(D_i - S_i)$ , if  $p_i > 0; h(D_i - S_i) = [0.01(D_i - S_i)]^+$ , if  $p_i = 0$ . The evolution of the token price for the three intervals  $T_1, T_2, T_3$  is shown for four different sets of initial prices (randomly chosen) for each of the six example test instances in Fig. 9. As can be seen, in all scenarios the prices converge to the same values for the three intervals, and all the market clearing conditions are satisfied for these prices. The equilibrium flows were also found to converge to the same values for all four initial sets of prices in all six scenarios.

Next, we examine convergence of the second price adjustment scheme (Eq. (18)) in Fig. 10. Note that here the price update rule is used within a interval halving scheme wherein the size of interval  $\Delta_p$  is halved if there is a price increase followed by a price decrease in successive iterations (or vice versa). Once again, we observe that in all six scenarios, the four different initial prices converge to the same unique prices for all three intervals, and these are identical to the prices obtained under the first price adjustment scheme. In a similar manner, the price adjustment schemes were verified to converge in all 324 test scenarios.

### Appendix B. Proof of Proposition 1

**Proof.** Let  $\Omega_p(\tau^*)$  and  $\Omega_q(\delta^*, p^*)$  denote the optimum welfare attained by the price and quantity instruments respectively, where  $\tau^*$  is the optimum vector of tolls (assume  $\tau^* > 0$  without loss of generality),  $\delta^*$  is a vector of optimum number of tokens required for each time interval, and  $p^*$  is the vector of market clearing prices. Note that for simplicity (w.l.o.g), we do not adopt the normalization of  $\delta_i = 1$  and instead assume that  $M$  is fixed arbitrarily, and the regulator optimizes  $\delta$ . Further, let the optimum flows obtained under the price instrument be denoted by  $X^p = (X_i^p, i = 0 \dots I)$ . We wish to show that  $\Omega_p(\tau^*) = \Omega_q(\delta^*, p^*)$ .

Since  $\tau^*$  is the optimum toll vector, we have

$$\Omega(\tau^*) > \Omega(\tau) \quad \forall \tau \neq \tau^*. \tag{31}$$

First, assume that  $p^*$  is given exogenously. Clearly, if we set  $\delta_i^* = \tau_i^*/p_i^*, i = 1 \dots I$ , the flows under the quantity instrument satisfy  $X^q(\delta^*, p^*) = X^p$ , and hence,  $\Omega_q(\delta^*, p^*) = \Omega_p(\tau^*)$ . The prices  $p^*$  can be determined from the market clearing conditions in Eq. (14), which imply that (note that since  $\tau^* > 0$ , it must be the case that  $\delta_i > M, i = 1 \dots I$  and hence,  $p^* > 0$ ),

$$Q_i X_0^p M = X_i^p (\delta_i - M), \forall i \in I. \tag{32}$$

Substituting  $\delta_i^* = \tau_i^*/p_i^*, i = 1 \dots I$ , we obtain the market clearing prices,

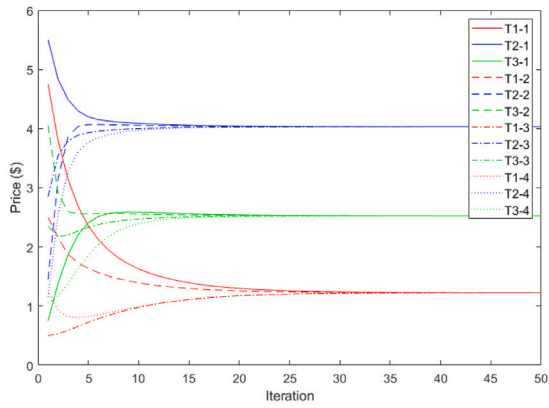
$$p_i^* = \frac{\tau_i^*}{M \left( 1 + \frac{Q_i X_0^p}{X_i^p} \right)}, \forall i \in I. \tag{33}$$

Note that  $M$  can be set arbitrarily; the prices will adjust accordingly. Thus, to summarize, by setting  $\delta_i^* = \tau_i^*/p_i^*, i = 1 \dots I$ , where  $p_i^*$  is given by Eq. (33), we have  $\Omega_q(\delta^*, p^*) = \Omega_p(\tau^*)$ .

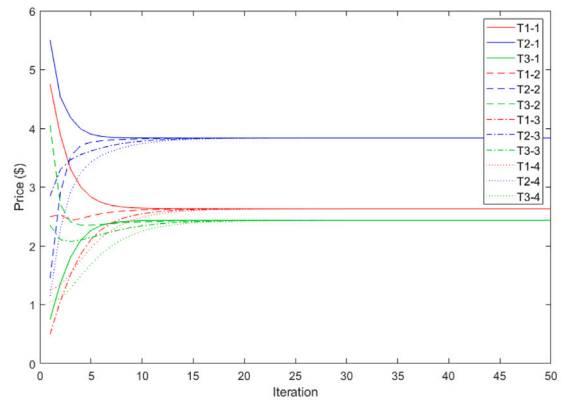
Now, assume that there exists  $\delta = \delta'$  and  $p = \bar{p}$  such that  $\Omega_q(\delta', \bar{p}) > \Omega_p(\tau^*)$ . Then, setting  $\tau'_i = \bar{p}_i \delta'_i (i = 1 \dots I)$ , we have  $\Omega_p(\tau') > \Omega_p(\tau^*)$ . This contradicts Eq. (31) and hence, the result follows. ■

### Appendix C. Note on Conjecture 1

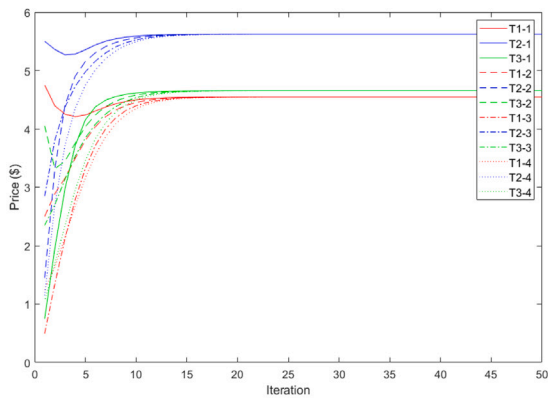
The intuition for [Conjecture 1](#) is as follows. The case  $\bar{\mu} = 0$  corresponds to a deterministic selling model. In this case, there is perfect competition between the players (users) so that all prices have to be the same, since otherwise an arbitrage opportunity would exist. Thus, no token will be sold in a given interval if its price is lower than that in another interval (with an active market, i.e., a period where tokens are bought or sold). On the other hand, all users would be willing to sell their tokens if the price in a period were higher than the price in any other interval. As a consequence all prices are the same on any active period.  $\bar{\mu} = 0$  would give rise to strong reaction as soon as a price is infinitesimally larger or smaller than the others and for this reason we introduce a smooth price adjustment mechanism. Introducing the Logit model to guarantee the existence of an equilibrium is a well accepted methodology in several fields, especially in economics (see [Anderson et al. \(1992\)](#)).



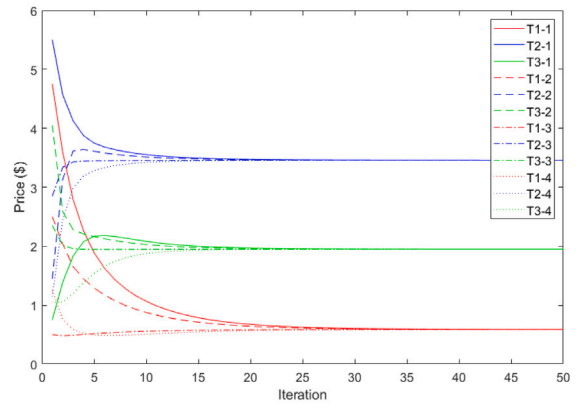
(a) Scenario 1:  $N = 1550, \beta = 3$



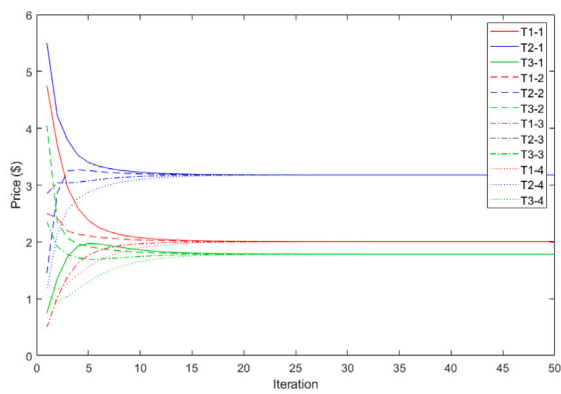
(b) Scenario 2:  $N = 1550, \beta = 4$



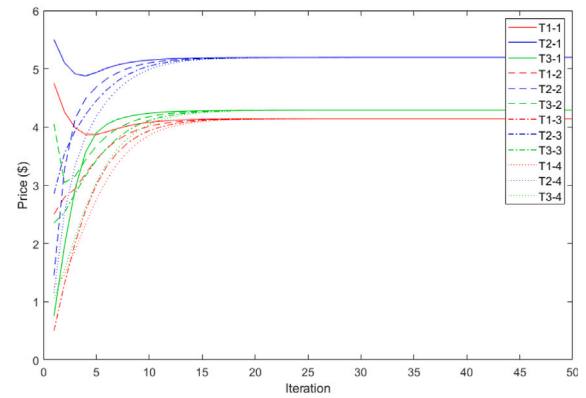
(c) Scenario 3:  $N = 1550, \beta = 6$



(d) Scenario 4:  $N = 1700, \beta = 3$

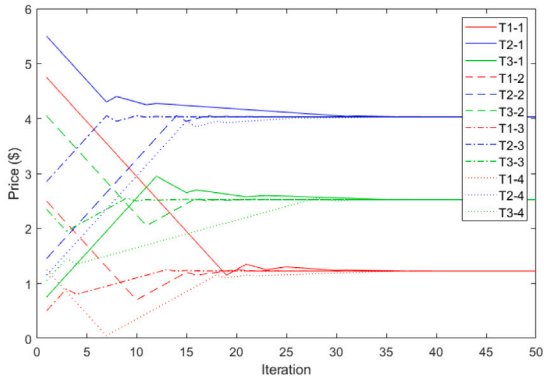


(e) Scenario 5:  $N = 1700, \beta = 4$

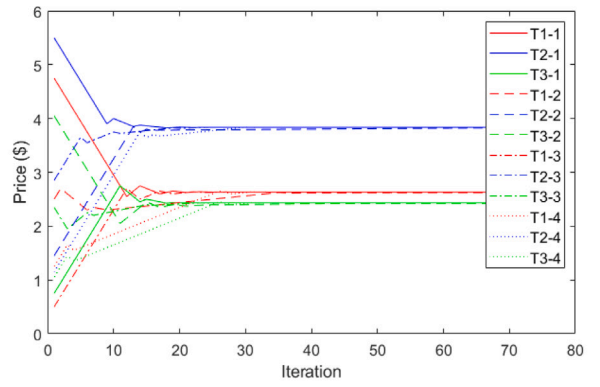


(f) Scenario 6:  $N = 1700, \beta = 6$

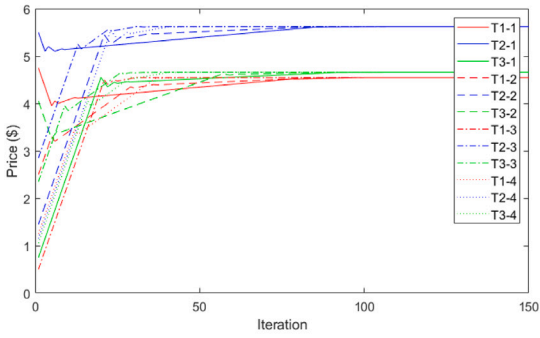
Fig. 9. Convergence of price adjustment scheme (Eq. (17)).



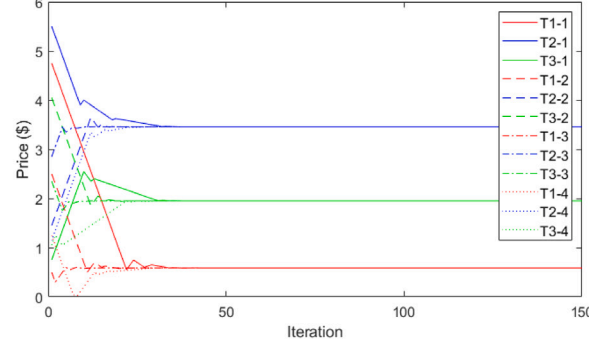
(a) Scenario 1:  $N = 1550, \beta = 3$



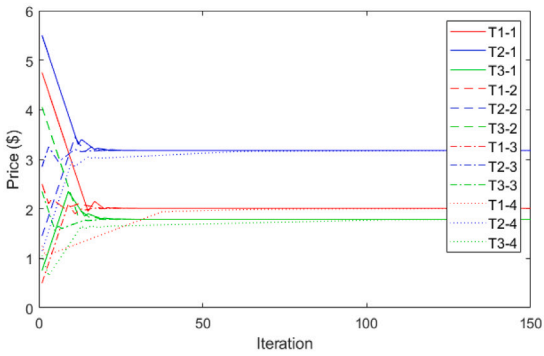
(b) Scenario 2:  $N = 1550, \beta = 4$



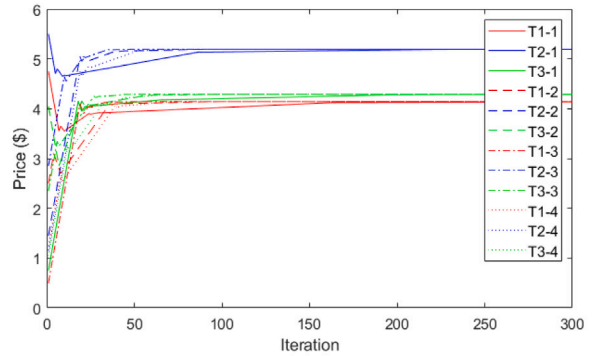
(c) Scenario 3:  $N = 1550, \beta = 6$



(d) Scenario 4:  $N = 1700, \beta = 3$



(e) Scenario 5:  $N = 1700, \beta = 4$



(f) Scenario 6:  $N = 1700, \beta = 6$

Fig. 10. Convergence of price adjustment scheme (Eq. (18)).

### Appendix D. Glossary

Notation	
Symbol	Description
$B_i^n$	Alternative specific benefit for individual $n$ in time interval $T_i$ ( $i = 0 \dots I$ )
$C_i$	Capacity in time interval $T_i$ ( $i = 1 \dots I$ )
$D_i$	Demand for tokens in time interval $T_i, i \in \tilde{I}$
$I$	Set of time intervals ( $T_i, i = 0 \dots I$ )
$\tilde{I}$	Subset of time intervals $T_i$ where $\delta_i > M$
$M$	Token endowment per traveler

$M^{sk}$	Token endowment per traveler on day $sk$ ( $k = 1, 2$ )
$N$	Number of travelers
$p_i$	Token market price in time interval $T_i$ ( $i = 1 \dots I$ )
$\bar{p}_i$	Perceived token market price for time interval $T_i$ ( $i = 1 \dots I$ )
$p_i^{sk}$	Token market price in time interval $T_i$ ( $i = 1 \dots I$ ) on day $sk$ ( $k = 1, 2$ )
$Q_i$	Probability of selling in time interval $T_i, i \in \tilde{I}$
$sk$	State of nature or day ( $k = 1, 2$ )
$S_i$	Supply of tokens in time interval $T_i, i \in \tilde{I}$
$t_i^{FF}$	Free flow travel time in interval $T_i$ ( $i = 1 \dots I$ )
$t_i$	Congested travel time in interval $T_i$ ( $i = 1 \dots I$ )
$T_i$	Time interval $i$ ( $i = 0 \dots I$ )
$U^n(T_i)$	Utility of time interval $i$ for individual $n$
$V^n(T_i)$	Systematic utility of time interval $i$ for individual $n$
$V_{sk}^n(T_i^{sk})$	Systematic utility of time interval $i$ on day $sk$ for individual $n$
$X_i$	Flow in time interval $T_i$ ( $i = 1 \dots I$ )
$X_i^{sk}$	Flow in time interval $T_i$ ( $i = 1 \dots I$ ) on day $sk$ ( $k = 1, 2$ )
$\alpha^n$	Value of time of individual $n$
$\alpha_i$	BPR function parameter for interval $T_i$ ( $i = 1 \dots I$ )
$\beta_i$	BPR function parameter for interval $T_i$ ( $i = 1 \dots I$ )
$\delta_i$	Toll in tokens for time interval $T_i$ ( $i = 1 \dots I$ )
$\epsilon_i$	Error term in utility for time interval $T_i$
$\mu_n$	Scale parameter of individual $n$ (mobility decision)
$\bar{\mu}$	Scale parameter of selling model
$\Omega_P$	Optimum welfare of price instrument
$\Omega_Q$	Optimum welfare of quantity instrument
$\tau_i$	Toll in dollars for time interval $T_i$ ( $i = 1 \dots I$ )

## References

- Akamatsu, T., Wada, K., 2017. Tradable network permits: A new scheme for the most efficient use of network capacity. *Transp. Res. C* 79, 178–195.
- Anderson, S.P., de Palma, A., Thisse, J.F., 1992. *Discrete Choice Theory of Product Differentiation*. MIT Press.
- Bao, Y., Verhoef, E.T., Koster, P., 2019. Regulating dynamic congestion externalities with tradable credit schemes: Does a unique equilibrium exist? *Transp. Res. B* 127, 225–236.
- Brands, D.K., Verhoef, E.T., Knockaert, J., Koster, P.R., 2020. Tradable permits to manage urban mobility: market design and experimental implementation. *Transp. Res. Part A: Policy Pract.* 137, 34–46.
- Chen, S., Seshadri, R., Azevedo, C.L., Akkinipally, A.P., Liu, R., Araldo, A., Jiang, Y., Ben-Akiva, M.E., 2021. Analysis and design of markets for tradable mobility credit schemes. *arXiv preprint arXiv:2101.00669*.
- Chen, A., Yang, H., Lo, H.K., Tang, W.H., 2002. Capacity reliability of a road network: an assessment methodology and numerical results. *Transp. Res. B* 36, 225–252.
- Cirillo, C., Axhausen, K.W., 2006. Evidence on the distribution of values of travel time savings from a six-week diary. *Transp. Res. Part A: Policy Pract.* 40, 444–457.
- De Jong, G., Daly, A., Pieters, M., Van der Hoorn, T., 2007. The logsum as an evaluation measure: Review of the literature and new results. *Transp. Res. Part A: Policy Pract.* 41, 874–889.
- de Palma, A., Lindsey, R., 2011. Traffic congestion pricing methodologies and technologies. *Transp. Res. C* 19, 1377–1399.
- de Palma, A., Lindsey, R., 2020. Tradable permit schemes for congestible facilities with uncertain supply and demand. *Econ. Transp.* 21, 100149.
- de Palma, A., Marchal, F., Nesterov, Y., 1997. Metropolis: Modular system for dynamic traffic simulation. *Transp. Res. Rec.* 1607, 178–184.
- de Palma, A., Proost, S., Seshadri, R., Ben-Akiva, M., 2018. Congestion tolling-dollars versus tokens: A comparative analysis. *Transp. Res. B* 108, 261–280.
- Deep, K., Singh, K.P., Kansal, M.L., Mohan, C., 2009. A real coded genetic algorithm for solving integer and mixed integer optimization problems. *Appl. Math. Comput.* 212, 505–518.
- Delle Site, P., de Palma, A., Kilani, K., et al., 2021. Consumers' Welfare and Compensating Variation: Survey and Mode Choice Application. Technical Report, THEMA (Théorie Economique, Modélisation et Applications), CY Cergy Paris Université.
- Dogterom, N., Ettema, D., Dijkstra, M., 2017. Tradable credits for managing car travel: A review of empirical research and relevant behavioural approaches. *Transp. Res. B* 37, 322–343.
- Fan, W., Jiang, X., 2013. Tradable mobility permits in roadway capacity allocation: Review and appraisal. *Transp. Policy* 30, 132–142.
- Goddard, H.C., 1997. Using tradeable permits to achieve sustainability in the world's large cities: policy design issues and efficiency conditions for controlling vehicle emissions, congestion and urban decentralization with an application to Mexico city. *Environ. Resour. Econ.* 10, 63–99.
- Grant-Muller, S., Xu, M., 2014. The role of tradable credit schemes in road traffic congestion management. *Transp. Res. B* 57, 128–149.
- He, F., Yin, Y., Shirmohammadi, N., Nie, Y.M., 2013. Tradable credit schemes on networks with mixed equilibrium behaviors. *Transp. Res. B* 57, 47–65.
- Hess, S., Bierlaire, M., Polak, J.W., 2005. Estimation of value of travel-time savings using mixed logit models. *Transp. Res. Part A: Policy Pract.* 39, 221–236.
- Jaensirisak, S., Wardman, M., May, A., 2005. Explaining variations in public acceptability of road pricing schemes. *J. Transp. Econ. Policy (JTPEP)* 39, 127–154.
- Kalmanje, S., Kockelman, K.M., 2004. Credit-based congestion pricing: Travel, land value, and welfare impacts. *Transp. Res. Rec.* 1864, 45–53.
- Kockelman, K.M., Kalmanje, S., 2005. Credit-based congestion pricing: A policy proposal and the public's response. *Transp. Res. Part A: Policy Pract.* 39, 671–690.
- Konishi, H., 2004. Uniqueness of user equilibrium in transportation networks with heterogeneous commuters. *Transp. Sci.* 38, 315–330.
- Laffont, J.J., 1977. More on prices vs. quantities. *Rev. Econom. Stud.* 44, 177–182.
- Lamotte, R., Geroliminis, N., 2021. Monotonicity in the trip scheduling problem. *Transp. Res. B* 146, 14–25.
- Lamotte, R., de Palma, A., Geroliminis, N., 2022. Impacts of metering-based dynamic priority schemes. *Transp. Sci.* 56, 358–380.



- Lindsey, R., 2004. Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes. *Transp. Sci.* 38, 293–314.
- Lindsey, R., de Palma, A., Rezaeinia, P., 2022. Tolls vs tradable permits for managing travel on a bimodal congested network with variable capacities and demands. Available at SSRN 4066992.
- Liu, R., Chen, S., Jiang, Y., Seshadri, R., Ben-Akiva, M.E., Azevedo, C.L., 2020. Managing network congestion with a tradable credit scheme: a trip-based mfd approach. *arXiv preprint arXiv:2009.06965*.
- Liu, Y., Nie, Y.M., 2017. A credit-based congestion management scheme in general two-mode networks with multiclass users. *Netw. Spat. Econ.* 17, 681–711.
- Meng, Q., Lee, D.H., Yang, H., Huang, H.J., 2004. Transportation network optimization problems with stochastic user equilibrium constraints. *Transp. Res. Rec., J. Transp. Res. Board* 1882, 113–119.
- Miralinaghi, M., Peeta, S., 2016. Multi-period equilibrium modeling planning framework for tradable credit schemes. *Transp. Res. Part E: Logist. Transp. Rev.* 93, 177–198.
- Miralinaghi, M., Peeta, S., 2019. Promoting zero-emissions vehicles using robust multi-period tradable credit scheme. *Transp. Res. Part D: Transp. Environ.* 75, 265–285.
- Nakayama, S., Connors, R., 2014. A quasi-dynamic assignment model that guarantees unique network equilibrium. *Transportmetrica A: Transp. Sci.* 10, 669–692.
- Nie, Y., 2012. Transaction costs and tradable mobility permits. *Transp. Res. Part B* 46, 189–203.
- Nie, Y.M., 2015. A new tradable credit scheme for the morning commute problem. *Netw. Spat. Econ.* 15, 719–741.
- Nie, Y.M., Yin, Y., 2013. Managing rush hour travel choices with tradable credit scheme. *Transp. Res. B* 50, 1–19.
- Prato, C.G., Rasmussen, T.K., Nielsen, O.A., 2014. Estimating value of congestion and of reliability from observation of route choice behavior of car drivers. *Transp. Res. Rec.* 2412, 20–27.
- Raux, C., 2004. The use of transferable permits in transport policy. *Transp. Res. Part D: Transp. Environ.* 9, 185–197.
- Rezaeinia, P., de Palma, A., Lindsey, R., 2021. Traffic Congestion Control: Tradable Permits Vs Road Tolls. Technical Report, THEMA (Théorie Economique, Modélisation et Applications), CY Cergy Paris Université.
- Santos, G., Verhoef, E., 2011. Road congestion pricing. In: *A Handbook of Transport Economics*. Edward Elgar Publishing.
- Shirmohammadi, N., Zangui, M., Yin, Y., Nie, Y., 2013. Analysis and design of tradable credit schemes under uncertainty. *Transp. Res. Rec.* 2333, 27–36.
- Small, K.A., Rosen, H.S., 1981. Applied welfare economics with discrete choice models. *Econometrica* 10, 5–130.
- Tian, L.J., Yang, H., Huang, H.J., 2013. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. *Transp. Res. Part E: Logist. Transp. Rev.* 54, 1–13.
- Tranroy, A., 2011. Equity dimensions of transport policies. In: *Handbook of Transport Economics*. Edward Elgar, Cheltenham, UK, pp. 625–649.
- Verhoef, E., Nijkamp, P., Rietveld, P., 1997. Tradeable permits: their potential in the regulation of road transport externalities. *Environ. Plan. B: Plann. Des.* 24, 527–548.
- Wang, G., Gao, Z., Xu, M., 2014a. An mpec formulation and its cutting constraint algorithm for continuous network design problem with multi-user classes. *Appl. Math. Model.* 38, 1846–1858.
- Wang, G., Gao, Z., Xu, M., Sun, H., 2014b. Models and a relaxation algorithm for continuous network design problem with a tradable credit scheme and equity constraints. *Comput. Oper. Res.* 41, 252–261.
- Wang, X., Yang, H., 2012. Bisection-based trial-and-error implementation of marginal cost pricing and tradable credit scheme. *Transp. Res. B* 46, 1085–1096.
- Wang, X., Yang, H., Han, D., Liu, W., 2014c. Trial and error method for optimal tradable credit schemes: The network case. *J. Adv. Transp.* 48, 685–700.
- Wang, X., Yang, H., Zhu, D., Li, C., 2012. Tradable travel credits for congestion management with heterogeneous users. *Transp. Res. Part E: Logist. Transp. Rev.* 48, 426–437.
- Weitzman, M.L., 1974. Prices vs. quantities. *Rev. Econom. Stud.* 41, 477–491.
- Wu, D., Yin, Y., Lawphongpanich, S., Yang, H., 2012. Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. *Transp. Res. B* 46, 1273–1287.
- Xiao, F., Qian, Z.S., Zhang, H.M., 2013. Managing bottleneck congestion with tradable credits. *Transp. Res. B* 56, 1–14.
- Yang, H., Wang, X., 2011. Managing network mobility with tradable credits. *Transp. Res. B* 45, 580–594.
- Ye, H., Yang, H., 2013. Continuous price and flow dynamics of tradable mobility credits. *Procedia-Soc. Behav. Sci.* 80, 61–78.
- Zhang, X., Yang, H., 2004. The optimal cordon-based network congestion pricing problem. *Transp. Res. B* 38, 517–537.
- Zhu, D.L., Yang, H., Li, C.M., Wang, X.L., 2015. Properties of the multiclass traffic network equilibria under a tradable credit scheme. *Transp. Sci.* 49, 519–534.