Preference heterogeneity and congestion pricing: the two route case revisited☆

Paul Koster[a,b,*], Erik Verhoef[a,b], Simon Shepherd[c], David Watling[c]

[a]*Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands*
[b]*Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands*
[c]*Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK*

Abstract

This paper studies first-best and second-best congestion pricing in the presence of unobserved and observed preference heterogeneity using a stylised stochastic user equilibrium choice model. Travellers choose between multiple alternatives, have heterogeneous values of travel times, and may differ in their valuation of variety. We derive first-best and second-best tolls taking into account how the overall network demand responds to expected generalized prices, including tolls. For second-best pricing, we show that with homogeneous values of times the welfare losses of second-best pricing are smaller when route choice is probabilistic than when route choice is deterministic. Furthermore, we find that with heterogeneous values of times and benefits of variety, uniform second-best tolls and group-differentiated tolls can be very close, implying potentially low welfare losses from the inability to differentiate tolls. Finally, we show that there are cases where all groups benefit from second-best congestion pricing, but that these cases are likely to be politically unacceptable because tolls are then higher for low income groups.

Keywords: Stochastic user equilibrium, Second-best congestion pricing, Preference heterogeneity, Scale heterogeneity, Probabilistic choice

---

☆Erik Verhoef and Paul Koster gratefully acknowledge the financial support of the Advanced ERC Grant OPTION # 246969. David Watling and Simon Shepherd acknowledge the support of the EPSRC in funding the research as part of the project "Competitive Cities : The network and long-term impacts of fiscal management of transport demand" grant number EP/H021345/1." We benefited from comments from seminar participants at seminars in Amsterdam and Leeds.

*Corresponding author. Fax +31 20 5986004, phone +31 20 5982369.

Email addresses: `p.r.koster@vu.nl` (Paul Koster), `e.t.verhoef@vu.nl` (Erik Verhoef), `S.P.Shepherd@its.leeds.ac.uk` (Simon Shepherd), `D.P.Watling@its.leeds.ac.uk` (David Watling)

1. Introduction

The clustering of human activities in time and space results in substantial social costs of congestion. For the year 2050, it is expected that 66 percent of the world population will live in urban areas (United Nations, 2014), and with this ongoing increase of urbanization, levels of congestion are expected to increase as well. Nash (2003) estimates congestion costs for European countries at about 1% of GDP, meaning that potential welfare improvements from the regulation of congestion externalities can be substantial. Since the seminal work of Pigou (1920) economists have argued that the price of travellers' trips does not correspond to the marginal social costs because a driver does not take into account that (s)he raises the travel time costs of other travellers' on the road (see Walters (1961) for an early contribution). Therefore congestion pricing has long been advocated as a viable solution, but political and societal opposition has limited its implementation.

Unlike what is assumed in the earliest contributions to the road pricing literature, researchers cannot observe all determinants of choice. Stochastic User Equilibrium (SUE) models are therefore widely employed, for example to study pricing and location decisions of firms (Anderson et al. (1992)), households' location choices (Bayer and Timmins (2007)), and route choices of travellers (Daganzo and Sheffi (1977)). Instead of considering purely deterministic trade-offs, the utility of alternatives is assumed to depend on a deterministic part and an unobserved part, that might vary over individuals as well as over choice occasions. Individuals' unobserved preferences for routes or modes result in "benefits of variety": an increased number of routes or modes will raise the expected utility of travelling because different alternatives may be appealing to different subsets of consumers. The variety benefits can be included in the welfare function using an entropy term. For example, Erlander (1977), Fisk (1980), Miyagi (1986) and Anderson et al. (1988) showed the connection between the logit model of discrete choices and the benefits of variety: when alternatives have exactly the same deterministic utility (in equilibrium), and hence the same choice probabilities, the benefit of variety is maximized. This corresponds to the intuitive notion that additional alternatives that are (almost) unused in equilibrium hardly increase variety benefits.

1.1. Contribution

This paper shows analytically and numerically how observed and unobserved preference heterogeneity in SUE impacts first-best and second-best congestion pricing policies. We include both heterogeneity in the deterministic part of utility, for example caused by the fact that travellers value travel time differently (Small (2012)), and in the unobserved part of utility by allowing for group specific substitution parameters. Because congestion taxes may impact the benefits of variety and the deterministic part of utility of different groups differently, including preference heterogeneity is of key importance to provide policy makers information about the distributional impacts of congestion pricing. Furthermore, the welfare benefits of congestion pricing may be higher when differentiation of congestion taxes between groups is feasible. The main body of this paper looks at a stylised two-route case to enhance economic interpretation of first-best and second-best congestion tolling with choices governed by random utility maximization. It extends the two-route deterministic user equilibrium

(DUE) models of Verhoef et al. (1996) and Small and Yan (2001) to account for the valuation of route variety and an arbitrary number of groups with distinct preferences. Our stylised analytical approach can also be applied in the analysis of taxation of other externalities in the presence of variety benefits and heterogeneous preferences. Extensions to an arbitrary number of alternatives are provided in the appendices.

Several earlier studies have studied congestion tolling in SUE network models (see Yang (1999); Yang and Huang (2004); Maher et al. (2005); Huang and Li (2007)) and have analysed congestion pricing with heterogeneous preferences (Arnott et al. (1994); Verhoef et al. (1995); Small and Yan (2001); Verhoef and Small (2004); Mahmassani et al. (2005); Lu et al. (2006); Zhang et al. (2008); Clark et al. (2009); Jiang et al. (2011); Sumalee and Xu (2011); van den Berg and Verhoef (2011a,b, 2013)). However, the most likely realistic combination of price-sensitivity of demand, heterogeneity in valuations of travel time, and benefits of variety has not been studied in a stylised network before. As we accommodate several sources of preference heterogeneity in a fairly general way, it can inspire future analytical research on taxation of externalities in networks in transportation and beyond.

1.2. Structure of the paper and main findings

After introducing the behavioural model in Section 2, Section 3 introduces first-best congestion pricing using a probabilistic SUE model. First, we derive analytical expressions for first-best congestion tolling with homogeneous values of travel time (VOT) and valuation of variety (see Section 3.1). We show that probabilistic choice has no impact on the first-best toll rules, when compared with the Pigouvian toll rules of the Deterministic User Equilibrium (DUE) model. However, for asymmetric route costs, the *levels* of these first-best tolls may still differ for SUE and DUE, despite the equality of the toll rule, because SUE and DUE equilibrium flows are different and therefore so are the marginal external costs. These results also hold for an arbitrary number of alternatives.

Second, we derive first-best congestion tolls in the presence of heterogeneous values of time and benefits of variety (see Section 3.2). Our model thus allows for scale heterogeneity, meaning that the benefits of variety may differ between groups. The DUE model with two groups of Small and Yan (2001) is a limiting case of our model. We assume a finite number of groups, with each group having a different valuation of travel time, and valuation of route variety.[1] When first-best congestion tolls are group-specific, the SUE tolls have the same analytical form as the DUE tolls. The marginal expressions do depend on the group-specific valuations of travel times, but are independent of the benefits of variety. But again, the SUE toll levels may be different when route costs are asymmetric, because equilibrium aggregate usage levels are. The uniform first-best toll we find is equal to the group-specific first-best toll, because we assume that each traveller raises congestion by the same amount. This

[1] Although a continuous distribution may be even more realistic, compared to the case with homogeneous preferences, a discrete distribution of VOTs and scale parameters strongly increases the empirical plausibility of the model and can be connected to empirical applications that seek to estimate preference heterogeneity. For example, it is well known that VOTs of travellers may be different because of variations in job and other characteristics (see Small (2012) for a recent review on heterogeneity in VOTs).

result also holds for an arbitrary number of alternatives as shown in Appendix A.

Section 4 studies second-best congestion pricing using a SUE modelling framework. First, we derive a second-best toll with homogeneous VOTs and benefits of variety, which has the deterministic second-best toll of Verhoef et al. (1996) as a limiting case when benefits of variety vanish (see Section 4.1). Here we find that the toll rule of the DUE model and the SUE model diverge. This is because the second-best toll corrects for the spillovers on the untolled route. The substitution effect to the untolled route depends on the relative size of the random idiosyncratic part of utility in the total utility. Lower benefits of variety arise when route choices are more deterministic and this naturally will lead to a stronger behavioural response to the toll, in turn leading to a higher diversion of travellers onto the untolled route when second-best pricing is employed and therefore a larger downward adjustment of the second-best toll to limit this spill-over. Because drivers are less responsive to tolls in the SUE model, these second-best tolls will be higher: the spillover effect on the untolled route is smaller. An extension to an arbitrary number of alternatives is provided in Appendix B.

Second, we derive a group-specific second-best toll rule with preference heterogeneity in the systematic and the random part of utility (see Section 4.2). The level of this second-best toll depends on the benefits of variety of the different groups. We find that there are cases where both groups benefit from SB pricing, but that it is likely to be politically unacceptable to implement this policy because the low value of time group faces higher tolls. For (non-differentiated) uniform second-best tolls, we were not able to derive an analytical closed-form expression, and therefore this case is analysed numerically. An extension to an arbitrary number of alternatives is provided in Appendix C.

Section 5 confirms the analytical expressions for first and second-best tolls and gives additional insights on the role of variety benefits and the distributional impacts of congestion tolling. We assume that there are two routes and two groups: one with high VOTs, and one with low VOTs. With low benefits of variety, there may be a toll differentiated equilibrium, where the high VOT group uses the tolled road and the low VOTs group uses the untolled route. However, when the benefits of variety increase, this separation disappears, due to unobserved route preferences becoming more important in route choice, and a pooled equilibrium is optimal. The extent to which it is beneficial to differentiate the road taxes in order to accommodate the needs of distinct groups therefore crucially depends on the valuation of variety and the heterogeneity in the value of travel times. Higher valuation of variety and lower heterogeneity in travel time valuations decrease the likelihood of having a separated optimal equilibrium. This result nuances earlier findings on toll differentiation based on deterministic models (Verhoef and Small, 2004; Small and Yan, 2001). Furthermore, the distributional impacts of SB tolling strongly depend on the benefits of variety. With deterministic route choice, the high VOT group benefits from SB tolling and the low VOT group loses. When route choice is more probabilistic the outcome is more nuanced and there are cases where both the low VOT and the high VOT group benefit from SB congestion tolling. Finally, Section 6 discusses extensions to more general networks and Section 7 concludes.

2. The random utility framework

We consider choice behaviour between two congestible facilities where choice is governed by random utility maximization. To fix ideas, we cast the analysis in terms of route choice on a congestible road network. Travellers choose their route on the basis of random utility maximization. The random utility function of a randomly sampled individual n belonging to group k, choosing route r from the set of two routes is:

$$U_{krn} = V_{kr} + \frac{1}{\theta_k}\epsilon_{krn}, r = 1, 2, k = 1...K \tag{1}$$

where U_{krn} depends on a deterministic part V_{kr} and a stochastic idiosyncratic route preference ϵ_{krn}, which is assumed to be identically and independently distributed. These unobserved preferences reflect route characteristics that affect route choice that are unobserved by the researcher. For example, one route might be closer to the child care or might be chosen because of travellers' habits. This leads to unobserved route preferences that differ over individuals and differ from one choice occasion to another. The scale parameter θ_k governs the relative importance of the unobserved idiosyncratic part of the utility in the total utility. When these idiosyncratic preferences are i.i.d. extreme value distributed, the route probabilities are given by the well known logit formula:[2]

$$P_{nkx} = \frac{\exp(\theta_k V_{kx})}{\exp(\theta_k V_{k1}) + \exp(\theta_k V_{k2})}, x = 1, 2, \forall k = 1...K. \tag{2}$$

Two limiting cases can be considered. First, the unobserved part of route utility may be very large ($\theta_k \to 0$), resulting in route choices that are independent of the deterministic part of utility. Choice probabilities then converge, in the limit, to 1/2. Second, the unobserved part of route utility may become very small ($\theta_k \to \infty$), resulting in a deterministic route choice model. Because the unobserved part of route utility is interpreted as individuals' unobserved preferences, there are benefits of variety in the sense that adding an additional alternative leads to higher expected utility (even when this alternative has lower systematic utility than all other available alternatives). This becomes more evident when we derive the expected utility, which is given by the expectation of equation (1) over all alternatives.[3]

$$EU_{nk} = \frac{1}{\theta_k} \ln\left[\exp(\theta_k V_{k1}) + \exp(\theta_k V_{k2})\right], \tag{3}$$

where one noteworthy feature of equation (3) is thus that having more routes to choose from is valued positively by travellers. The stochastic model has the deterministic model as a limiting case. When $\theta_k \to \infty$ in equation (3), the systematic utility differences between all used alternatives must become 0 in equilibrium if both routes are used, because otherwise all travellers will switch to the most attractive route.

[2] Luce, D. and Suppes (1965) attribute the proof to E.W. Holman and A.A.J. Marley. McFadden (1974) showed the reverse: logit probabilities necessarily imply i.i.d. extreme value distributed random utilities

[3] See Williams (1977) and Small and Rosen (1981) for theoretical contributions and de Jong et al. (2007) for a discussion on practical applications in transport.

3. First-best congestion pricing, two route case

3.1. Homogeneous preferences

We start our analysis with first-best congestion pricing in a stylised two-route setting. This model can be viewed as a probabilistic version of the model in Verhoef et al. (1996). We assume that there is only one group, meaning that all travellers are identical in terms of systematic utility and in terms of the distribution of unobserved utility. Therefore we can write $V_{kr} \equiv V_r$ and $\theta_k \equiv \theta$. The marginal benefit for entering the network is given by $D(N)$, where N is the total number of travellers. We assume that tolls and congestion costs enter systematic utility in an additive separable way, resulting in deterministic utilities for routes U and T of:

$$V_r = -\left(f_r + c_r(N_r)\right), r \in \{U, T\}, \tag{4}$$

where f_r is the toll on route r, and $c_r(N_r)$ is the travel cost for route r, which is increasing in the route flow N_r. For the two route case, equilibrium is implicitly defined by:

$$P_r = \frac{\exp(\theta V_r)}{\exp(\theta V_T) + \exp(\theta V_U)} = \frac{N_r}{N_U + N_T} = \frac{N_r}{N}, \tag{5}$$

meaning that equilibrium proportions can always be expressed by the number of drivers on the two routes. The social surplus S is given by the social benefits (the integral under the inverse demand curve, the toll revenues and the benefits of variety) minus the sum of total travel costs. The toll payments of the travellers are a money transfer to the regulator, and they drop out of the social surplus function (we assume zero transaction costs of taxation). In the absence of income effects, social surplus is given by (Fisk (1980) and Anderson et al. (1988)):

$$S = \int_0^N D(n)dn - N_T c_T(N_T) - N_U c_U(N_U) - \frac{1}{\theta}\left(N_T \ln\left[\frac{N_T}{N}\right] + N_U \ln\left[\frac{N_U}{N}\right]\right). \tag{6}$$

The first part of this equation captures the consumer surplus and the deterministic total user costs. The second part is always non-negative and captures the total benefits of variety for given route flows N_T and N_U. More precisely: here the individual route entropy $P_T \ln[P_T] + P_U \ln[P_U]$ is multiplied with the valuation of variety $\frac{1}{\theta}$, to obtain the benefits of variety per traveller, which is then multiplied with the total number of travellers to obtain the total benefits of variety. For given N_T and N_U, a smaller θ, and hence a higher randomness of route utility, will lead to higher benefits of variety. For the multinomial logit model, this relationship between this so-called "Shannon entropy" and the logit model has long been recognised (Erlander, 1977; Fisk, 1980; Miyagi, 1986; Anderson et al., 1988).

Entropy is higher when equilibrium route proportions are more alike. This is intuitive: the benefits of variety are higher when routes are used more equally in equilibrium. The total benefits of variety are fully expressed by the route proportions and the total number of travellers. Any change in the congestion costs or the toll on a route will only have an impact on these benefits via the equilibrium proportions. Furthermore, our model captures that increases in the benefits of variety will lead to additional overall demand. We now consider first-best congestion pricing by a welfare-maximizing regulator, setting a toll on route U and route T. The Lagrangian is given by:

$$\begin{aligned}\mathcal{L} = &\int_0^N D(n)dn - N_T c_T(N_T) - N_U c_U(N_U) - \frac{1}{\theta}\left(N_T \ln\left[\frac{N_T}{N}\right] + N_U \ln\left[\frac{N_U}{N}\right]\right) \\ &+ \lambda_T \left(D(N) - \frac{1}{\theta}\ln\left[\frac{N_T}{N}\right] - f_T - c_T(N_T)\right) + \lambda_U \left(D(N) - \frac{1}{\theta}\ln\left[\frac{N_U}{N}\right] - f_U - c_U(N_U)\right).\end{aligned} \tag{7}$$

The constraints govern equilibrium on both routes because travellers keep on entering the road up to the point where the marginal benefits $D(N) - \frac{1}{\theta}\ln\left[\frac{N_r}{N}\right]$ are equal to the generalised price of travelling: $f_r + c_r(N_r)$.[4] In order to find the first-best congestion tolls, the following first-order conditions need to be solved jointly:[5]

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial N_T} =& D(N) - c_T(N_T) - N_T c_T'(N_T) - \frac{1}{\theta}\ln\left[\frac{N_T}{N}\right] + \lambda_T \left(D'(N) - \frac{1}{\theta}\frac{N_U}{N_T N} - c_T'(N_T)\right) \\ &+ \lambda_U \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = 0.\end{aligned} \tag{8}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial N_U} =& D(N) - c_U(N_U) - N_U c_U'(N_U) - \frac{1}{\theta}\ln\left[\frac{N_U}{N}\right] + \lambda_T \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) \\ &+ \lambda_U \left(D'(N) - \frac{1}{\theta}\frac{N_T}{N_U N} - c_U'(N_U)\right) = 0.\end{aligned} \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial f_T} = -\lambda_T = 0. \tag{10}$$

$$\frac{\partial \mathcal{L}}{\partial f_U} = -\lambda_U = 0. \tag{11}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_T} = D(N) - f_T - c_T(N_T) - \frac{1}{\theta}\ln\left[\frac{N_T}{N}\right] = 0. \tag{12}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_U} = D(N) - f_U - c_U(N_U) - \frac{1}{\theta}\ln\left[\frac{N_U}{N}\right] = 0. \tag{13}$$

[4] The setup of equations (6) and (7) separates the overall demand response from the substitution between routes and has a clear advantage over an alternative model with a third "outside" alternative, because then θ governs both the elasticity of demand *and* the benefits of variety. The limiting case of $\theta \to \infty$ then results in a deterministic model with perfectly elastic demand. Our model is more general and has the nested logit with a degenerated nest for not travelling as a special case (see also Verboven (1996)).

[5] Here we use the derivatives of the entropy terms with respect to route demand which shows how benefits of variety change with route demands. The marginal entropy for route T is given by: $\frac{\partial}{\partial N_T}\left(N_T \ln\left[\frac{N_T}{N}\right] + N_U \ln\left[\frac{N_U}{N}\right]\right) = \ln\left[\frac{N_T}{N}\right]$, $\frac{\partial}{\partial N_U}\left(N_T \ln\left[\frac{N_T}{N}\right] + N_U \ln\left[\frac{N_U}{N}\right]\right) = \ln\left(\frac{N_U}{N}\right)$, $\frac{\partial}{\partial N_T}\ln\left[\frac{N_T}{N}\right] = \frac{N_U}{N_T N}$, $\frac{\partial}{\partial N_U}\ln\left[\frac{N_U}{N}\right] = \frac{N_T}{N_U N}$. The marginal entropy of route U is given by: $\frac{\partial}{\partial N_U}\ln\left[\frac{N_T}{N}\right] = \frac{\partial}{\partial N_T}\ln\left[\frac{N_U}{N}\right] = -\frac{1}{N}$.

Equations (10) and (11) show that the Lagrangian multipliers of both routes are 0 in the socially optimal equilibrium. This is intuitive, because these multipliers reflect the marginal change in social surplus for a unit change in the congestion toll on a route. In equilibrium, this should be 0, otherwise the tolls would be non-optimal by definition. This matches insights from deterministic models (Verhoef, 2002a). Substituting equations (10)-(13) in equations (8) and (9) we obtain:

$$\begin{aligned} f_T &= N_T c_T'(N_T), \\ f_U &= N_U c_U'(N_U). \end{aligned} \tag{14}$$

These first-best toll rules have the same form as the standard Pigouvian toll rules of the DUE model (this is also true for more than two alternatives). Tolls internalise marginal external cost to make people behave according to the social optimum, when acting in their own self-interest. In the probabilistic model we may not fully observe all the individual benefit components, but through the first-best tolls travellers are correctly taking into account all relevant aspects (their own costs and benefits, be it observable to the outsider or not, and the impact on other travellers), so they behave so as to maximise welfare. Even though the toll *rules* are the same for SUE and DUE, absolute toll *levels* may diverge when route costs are asymmetric. This asymmetric case is analysed in more detail in section 5. The extension to the multinomial case is straightforward and leads to the same toll expression.

3.2. Group-differentiated and common first-best tolls with heterogeneity in preferences

Next, we proceed with the analysis of first-best congestion pricing with heterogeneity in preferences. Assume that there are K distinct groups in the population. The inverse demand for travelling, the valuation of travel time, and the benefits of variety are assumed to be group-specific, and within each group route choice is governed by random utility maximization where the degree of substitution is group specific. The heterogeneity in travel time valuation enters the model via the deterministic route costs $c_{rk}(N_r)$ for group k, with N_r being the total number of travellers using route r. Heterogeneity in the benefits of variety is captured by having a group-specific scale parameter θ_k. Finally, heterogeneity in overall demand response is captured by having a group-specific marginal benefit function $D_k(N_k)$. Let N_{Tk} be the number of travellers of group k that use route T, and N_{Uk} the number of travellers of group k that use route U. The total number of travellers in a group is $N_k = N_{Tk} + N_{Uk}$. Here we have $N_T = \sum_{k=1}^{K} N_{Tk}$, $N_U = \sum_{k=1}^{K} N_{Uk}$ and $N_T + N_U = \sum_{k=1}^{K} N_k = N$. Because the number of groups can freely be chosen, our model can approximate any continuous distribution of preferences arbitrarily closely.

Deterministic route costs are determined by the total number of travellers on each route. To simplify matters, these travel costs are assumed to be equal up to a group-specific multiplicative term, implying that $c_{Tk}(N_T) = \alpha_k c_T(N_T)$ and $c_{Uk}(N_U) = \alpha_k c_U(N_U), \forall k = 1...K$. When $c_r(N_r)$ is interpreted as the travel time on route r, this model can be viewed as a model with travellers having different valuations of travel time α_k. To save notation we define $\bar{N_T}^{\alpha} = \sum_{k=1}^{K} \alpha_k N_{Tk}$ as the preference weighted number of travellers at route T, and $\bar{N_U}^{\alpha} = \sum_{k=1}^{K} \alpha_k N_{Uk}$ as the preference weighted number of travellers for route U. Equilibrium route probabilities for group k then satisfy $P_{Tk} = \frac{N_{Tk}}{N_k}$ and $P_{Uk} = \frac{N_{Uk}}{N_k}$. Deterministic costs $f_{rk} + c_{rk}(N_r)$ are governed by the total number of travellers on route r, whereas the total benefits of variety are given by the sum of the group-specific entropy multiplied by the number of travellers and the inverse of the group-specific scale parameter θ_k. The benefits of variety of group k are fully determined by the route proportions of group k. The interaction of the groups in the network is captured in the deterministic route costs. The Lagrangian is given by a straightforward extension of equation (7) to K groups:

$$\begin{aligned}\mathcal{L} = &\sum_{k=1}^{K} \int_0^{N_k} D_k(n_k)dn_k - \sum_{k=1}^{K} \alpha_k N_{Tk} c_T(N_T) - \sum_{k=1}^{K} \alpha_k N_{Uk} c_U(N_U) \\ &- \sum_{k=1}^{K} \frac{1}{\theta_k} \left(N_{Tk} \ln \left[\frac{N_{Tk}}{N_k} \right] + N_{Uk} \ln \left[\frac{N_{Uk}}{N_k} \right] \right) \\ &+ \sum_{k=1}^{K} \lambda_{Tk} \left(D_k(N_k) - f_{Tk} - \alpha_k c_T(N_T) - \frac{1}{\theta_k} \ln \left[\frac{N_{Tk}}{N_k} \right] \right) \\ &+ \sum_{k=1}^{K} \lambda_{Uk} \left(D_k(N_k) - f_{Uk} - \alpha_k c_U(N_U) - \frac{1}{\theta_k} \ln \left[\frac{N_{Uk}}{N_k} \right] \right). \end{aligned} \tag{15}$$

For all groups, the marginal willingness to pay $D_k(N_k)$ should be equal to the generalised price in equilibrium, resulting in $2K$ equilibrium constraints and corresponding Lagrangian multipliers. The system can be solved using the first-order conditions with respect to N_{Tl}, N_{Ul}, the Lagrange multipliers and the tolls for a chosen group l. In Appendix A we show that the group-specific first-best tolls with heterogeneous preferences are given by

$$\begin{aligned} f_{Tk} &= \bar{N_T}^{\alpha} c'_T(N_T), \\ f_{Uk} &= \bar{N_U}^{\alpha} c'_U(N_U). \end{aligned} \tag{16}$$

Marginal first-best tolls on the routes have therefore the same analytical form as the differentiated tolls of the deterministic model (this is also true for more than two alternatives). As with first-best tolling with homogeneous preferences, probabilistic choice only impacts the level of the toll level, not the marginal rules, via the equilibrium number of travellers on both routes. Furthermore, equation (16) shows that the first-best tolls are equal for all groups. This is because congestion is assumed to be anonymous: the change in external costs for an additional traveller is assumed to be the same for all groups. For external costs it does not matter to which group the traveller belongs, since travel time losses are assumed to increase with the same amount independent of the type of traveller.[6] This is of course not the case when groups have different impacts on travel times, as would be likely with trucks versus passenger cars (see for example de Palma et al. (2008) and Parry (2008)). The extension to the multinomial case is straightforward and leads to the same toll expression. Differentiated tolls will in general not be equal across groups for second-best congestion pricing, as we will show in the next section.

[6]This observation that congestion charges must be anonymous when drivers are observationally indistinguishable was made earlier by Arnott and Kraus (1998).

4. Second-best congestion pricing

4.1. Homogeneous preferences

In many cases first-best pricing is not feasible, and often not accepted, because travellers then do not have the opportunity to travel on an untolled route. Tolling one of the two routes (a form of second-best congestion pricing), may then be a viable alternative. In this section we analyse congestion pricing with probabilistic choice in the presence of an untolled alternative. The SUE model developed in this section is a probabilistic version of the DUE model of Verhoef et al. (1996), which has its roots in the pioneering DUE analyses of Marchand (1968) and Lévy-Lambert (1968). For this deterministic model, the substitution between the routes plays an important role for determining the second-best toll. Because in the SUE model this substitution is governed by the valuation of variety via θ, we expect that second-best tolls will depend on θ too. The systematic route utility for the tolled route is given by $V_T = -\left(f_T + c_T(N_T)\right)$, whereas for the untolled route it is given by $V_U = -c_U(N_U)$. Equilibrium is implicitly defined by equation (5), but the equilibrium conditions are different compared to the first-best case, because no toll is levied on route U. This has an effect on the generalised price, and on overall entropy, because overall demand is responsive to generalised price levels. The generalised price for route T is given by $f_T + c_T(N_T)$, whereas the generalised price of the untolled route is $c_U(N_U)$. Because tolls are a cost for the travellers and a benefit for the government, the toll revenues will not enter the total social surplus. Therefore the expression for the total social surplus (equation (6)) will not change. Because we have price-sensitive demand, travellers enter the road up to the point where for both routes the marginal benefits $D(N) - \frac{1}{\theta}\ln\left[\frac{N_r}{N}\right]$ are equal to the generalised route price. The Lagrangian is given by:

$$\begin{aligned}\mathcal{L} = &\int_0^N D(n)dn - N_T c_T(N_T) - N_U c_U(N_U) - \frac{1}{\theta}\left(N_T \ln\left[\frac{N_T}{N}\right] + N_U \ln\left[\frac{N_U}{N}\right]\right) \\ &+ \lambda_T\left(D(N) - f_T - c_T(N_T) - \frac{1}{\theta}\ln\left[\frac{N_T}{N}\right]\right) + \lambda_U\left(D(N) - c_U(N_U) - \frac{1}{\theta}\ln\left[\frac{N_U}{N}\right]\right).\end{aligned} \tag{17}$$

The second-best toll can be found by solving the following system of first-order conditions:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial N_T} = &D(N) - c_T(N_T) - N_T c_T'(N_T) - \frac{1}{\theta}\ln\left[\frac{N_T}{N}\right] \\ &+ \lambda_T\left(D'(N) - c_T'(N_T) - \frac{1}{\theta}\frac{N_U}{N_T N}\right) + \lambda_U\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = 0.\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial N_U} = &D(N) - c_U(N_U) - N_U c_U'(N_U) - \frac{1}{\theta}\ln\left[\frac{N_U}{N}\right] \\ &+ \lambda_T\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) + \lambda_U\left(D'(N) - c_U'(N_U) - \frac{1}{\theta}\frac{N_T}{N_U N}\right) = 0.\end{aligned} \tag{19}$$

$$\frac{\partial\mathcal{L}}{\partial f_T} = -\lambda_T = 0. \tag{20}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_T} = D(N) - f_T - c_T(N_T) - \frac{1}{\theta} \ln \left[\frac{N_T}{N} \right] = 0. \tag{21}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_U} = D(N) - c_U(N_U) - \frac{1}{\theta} \ln \left[\frac{N_U}{N} \right] = 0. \tag{22}$$

Using equations (18), (20) and (21) we obtain:

$$f_T = N_T c'_T(N_T) - \lambda_U \left(D'(N) + \frac{1}{\theta} \frac{1}{N} \right). \tag{23}$$

Using equations (22) and (20), we can solve equation (19) for λ_U:

$$\lambda_U = \frac{N_U c'_U(N_U)}{D'(N) - c'_U(N_U) - \frac{1}{\theta} \frac{N_T}{N_U N}}. \tag{24}$$

This Lagrangian multiplier is non-positive, implying that when a positive toll on route U would be feasible, this would result in a welfare increase because tolls enter the constraints in equation (17) negatively. This is in line with expectations, because raising a toll on route U will bring the equilibrium closer to the first-best optimum. Substituting equation (24) in equation (23) gives:

$$f_T = N_T c'_T(N_T) - N_U c'_U(N_U) \frac{-D'(N) - \frac{1}{\theta} \frac{1}{N}}{c'_U(N_U) - D'(N) - \frac{1}{\theta} \frac{N_T}{N_U N}}, \tag{25}$$

which becomes equal to the deterministic second-best rule in Verhoef et al. (1996) when $\theta \to \infty$. The first term in equation (25) is equal to the marginal external costs on route T in the second-best equilibrium. The second term is more complicated and corrects for the marginal costs of congestion caused by substitution to the untolled route when a toll is levied on route T. The marginal external costs on route U are multiplied by a fraction which depends on the sensitivity of the marginal benefits with respect to total demand, both systematically via $D'(N)$ and via the benefits of variety (via $\frac{1}{\theta}\frac{1}{N}$). It also depends on the ratio of the equilibrium number of travellers on both routes, the total number of travellers, and the slope of the congestion cost function of the untolled route. It shows that the second-best toll depends in a complicated way on the variety benefits via θ, since θ has a direct positive effect on the numerator and the denominator of the correction term, but also has an indirect effect on equation (25) via the equilibrium number of travellers. This last effect is the result of additional total demand when there are higher variety benefits in equilibrium.

A more detailed look at equation (25) shows that despite the additional terms due to stochastic route choice, it has a similar analytical structure as the toll rule for deterministic route choice, and can be written as $f_T = MEC_T + MEC_U \frac{\Delta N_U}{\Delta N_T}$, where MEC_r is the marginal external cost on route r. The marginal external costs on route U are weighted with a term $\frac{\Delta N_U}{\Delta N_T}$, which is the change in the equilibrium number of travellers on route U due to a change in the equilibrium number of travellers on route T. More specifically, equation (17) shows that the term $D'(N) + \frac{1}{\theta}\frac{1}{N}$ is the change in the constraint for route U due to a marginal change in N_T, whereas $-c'_U(N_U) + D'(N) + \frac{1}{\theta}\frac{N_T}{N_U N}$ is the change in the constraint for route U due to a marginal change in N_U. The ratio therefore gives $\frac{\Delta N_U}{\Delta N_T}$. As opposed to the first-best toll rules of equation (14), the toll rules of the DUE model of Verhoef et al. (1996) and our SUE model differ even for the case with symmetric route costs. Several limiting cases can be considered.

First, the second-best toll rule of the deterministic model is a limiting case of the stochastic model when its random component vanishes:

$$\lim_{\theta \to \infty} f_T = N_T c'_T(N_T) - N_U c'_U(N_U) \frac{-D'(N)}{c'_U(N_U) - D'(N)}. \tag{26}$$

This toll is isomorphic to the toll rule for the DUE model developed by Verhoef et al. (1996). The SUE model therefore has the DUE model as a limiting case, quite intuitively when $\theta \to \infty$ and idiosyncratic utility vanishes.

Second, for perfectly overall inelastic demand, $D'(N) \to -\infty$, the toll rule becomes equal to the difference in marginal external costs on the two routes:

$$\lim_{D'(N) \to -\infty} f_T = N_T c'_T(N_T) - N_U c'_U(N_U). \tag{27}$$

This toll rule is isomorphic to the toll rule of the DUE model with price-insensitive demand of Verhoef et al. (1996). Because there is no effect of tolling on the overall demand, the regulator only seeks to find the optimal route split. This produces the first-best outcome, so in itself it is no surprise that as with first-best tolls, the toll rules for the DUE and the SUE model become identical again. The level of the toll in equation (27) may well be different for DUE and SUE for asymmetric route costs, because θ has an effect on the optimal route split. Furthermore, equation (27) may be negative if in equilibrium the marginal external costs on route U are higher than the marginal external costs on route T. This means that travellers on route T would receive a subsidy instead of paying a toll.

Third, with perfectly elastic overall demand the toll rule becomes:

$$\lim_{D'(N) \to -0} f_T = N_T c'_T(N_T) - N_U c'_U(N_U) \frac{-\frac{1}{\theta}\frac{1}{N}}{c'_U(N_U) - \frac{1}{\theta}\frac{N_T}{N_U N}}. \tag{28}$$

This is clearly different from the DUE case (see (26)), where the second term vanishes as $-D'(N)$ becomes 0. For perfectly elastic overall demand the toll rule depends on θ, because the substitution between the routes depends on the love of variety, whereas the use of route U would be fully independent of f_T with deterministic route choice and perfectly elastic demand. The reason is that the toll on route T then cannot affect the use of route U, so there is no benefit from taking route U into account in the toll rule and it is used solely to optimize the use of route T alone. In the stochastic model, there remains an effect of the toll on the use of route U, also when $D' = 0$, and this is accounted for in the toll rule.

Fourth, if route U is uncongested, $c'_U(N_U) \to 0$, and the toll rule (25) reduces to:

$$\lim_{c'_U(N_U) \to 0} f_T = N_T c'_T(N_T). \tag{29}$$

which is again isomorphic to the toll rule in the deterministic model. The absence of congestion on route U then means that this route is optimally priced when it is not tolled. The regulator may therefore ignore route U, and needs only to consider the unconstrained optimal regulation of route T.

Finally, it turns out that both the deterministic model with perfect substitution and the stochastic model with imperfect substitution are part of a broader class of models with a user benefit function $B(N_T, N_U)$. The first derivative of this benefit function is the inverse demand and should be equal to the congestion costs plus the toll in equilibrium. The second-best congestion toll is then given by (Small et al. (2007), equation 4.47):

$$f_T = N_T c_T'(N_T) - N_U c_U'(N_U) \frac{-\frac{\partial^2 B}{\partial N_U \partial N_T}}{c_U'(N_U) - \frac{\partial^2 B}{\partial N_U^2}}. \tag{30}$$

For perfect substitutes we have $B(N_T, N_U) = \int_0^{N_T+N_U} D(n)dn$, whereas for stochastic route choice we observe from equation (6) that there are additional benefits of variety resulting in: $B(N_T, N_U) = \int_0^N D(n)dn - \frac{1}{\theta}\left(N_T \ln\left[\frac{N_T}{N}\right] + N_U \ln\left[\frac{N_U}{N}\right]\right)$. If we substitute the second order derivatives of this benefit function in equation (30), we arrive at equation (25). In Appendix B we provide an extension to the multiple routes case.

4.2. Group-specific second-best tolling with heterogeneous preferences

This section generalises the SUE model of the previous section by deriving group-specific second-best congestion tolls with heterogeneous travellers. We use a similar setup as in section 3.2 where K distinct groups have different preferences for congestion costs, benefits of variety, and inverse demand curves. The Lagrangian is given by:

$$\begin{aligned} \mathcal{L} = &\sum_{k=1}^{K} \int_0^{N_k} D_k(n_k)dn_k - \sum_{k=1}^{K} \alpha_k N_{Tk} c_T(N_T) - \sum_{k=1}^{K} \alpha_k N_{Uk} c_U(N_U) \\ &- \sum_{k=1}^{K} \frac{1}{\theta_k}\left(N_{Tk} \ln\left[\frac{N_{Tk}}{N_k}\right] + N_{Uk} \ln\left[\frac{N_{Uk}}{N_k}\right]\right) \\ &+ \sum_{k=1}^{K} \lambda_{Tk}\left(D_k(N_k) - f_{Tk} - \alpha_k c_T(N_T) - \frac{1}{\theta_k} \ln\left[\frac{N_{Tk}}{N_k}\right]\right) \\ &+ \sum_{k=1}^{K} \lambda_{Uk}\left(D_k(N_k) - \alpha_k c_U(N_U) - \frac{1}{\theta_k} \ln\left[\frac{N_{Uk}}{N_k}\right]\right) \end{aligned} \tag{31}$$

In Appendix B we first show that the group-specific Lagrangian multipliers of route U are non-positive, implying that the possibility to raise a positive toll on route U for any of the groups would increase welfare. Furthermore, we show that the second-best group-specific toll for group k is given by:

$$f_{Tk} = \bar{N_T}^{\alpha} c_T'(N_T) - \bar{N_U}^{\alpha} c_U'(N_U) \frac{-D_k'(N_k) - \frac{1}{\theta_k}\frac{1}{N_k}}{\alpha_k c_U'(N_U) - D_k'(N_k) + \frac{1}{\theta_k}\frac{N_{Tk}}{N_{Uk} N} + \phi_k}, \tag{32}$$

where

$$\phi_k = \sum_{\substack{l=1 \\ l \neq k}}^{K} \alpha_l c'_U(N_U) \frac{D'_k(N_k) - \frac{1}{\theta_k}\frac{N_{Tk}}{N_{Uk}N_k}}{D'_l(N_l) - \frac{1}{\theta_l}\frac{N_{Tl}}{N_{Ul}N_l}} > 0. \tag{33}$$

The first part in equation (32) is related to the external costs on the tolled route and is isomorphic to the first-best toll with heterogeneous preferences (equation (16)). The second part in equation (32) takes into account the substitution effect to the other route which is different for each group. Several limiting cases can be considered. First, when there is only one group, $\phi_k \to 0$, and equation (32) reduces to equation (25). Second, the DUE group-specific toll is a special case for which $\theta_k \to \infty, \forall k = 1...K$. This results in:

$$f_{Tk} = \bar{N_T}^{\alpha} c'_T(N_T) - \bar{N_U}^{\alpha} c'_U(N_U) \frac{-D'_k(N_k)}{\alpha_k c'_U(N_U) - D'_k(N_k) - c'_U(N_U)\sum_{\substack{l=1 \\ l \neq k}}^{K} \alpha_k \frac{D'_k(N_k)}{D'_l(N_l)}}. \tag{34}$$

When the slopes of the demand curves of all groups are equal we have $D'_l(N_l) \equiv D'_k(N_k) \equiv D'$ this reduces to:

$$f_{Tk} = \bar{N_T}^{\alpha} c'_T(N_T) - \bar{N_U}^{\alpha} c'_U(N_U) \frac{-D'}{c'_U(N_U)\sum_{l=1}^{K} \alpha_l - D'}. \tag{35}$$

This implies that the DUE model with equal slopes of the demand curves lead to common second-best tolls for all groups because $c'_U(N_U)\sum_{l=1}^{K} \alpha_k$ has the same value for all groups. For the SUE model with equal slopes of the inverse demand curves, tolls are still differentiated between groups, because the substitution effect to the untolled route does depend on the equilibrium number of travellers of each group on each route and the group-specific benefits of variety via θ_k. The extension to multiple routes is provided in Appendix C.

We were not able to derive analytical solutions for the common second-best toll case (undifferentiated between groups). The welfare for common second-best tolls will be lower than for the group-specific second-best tolls, because the inability to differentiate the tolls between user groups imposes an additional constraint. The Lagrangian problem is equivalent to (31) with $f_{Tk} \equiv f_T$. The next section will include numerical results for this case.

5. Numerical results

5.1. Introduction and calibration

Our numerical results build on those for the DUE model of Verhoef et al. (1996), who assumed linear inverse demand and linear congestion cost functions. We shall use the DUE case as a benchmark case against which we judge the implications of moving from a DUE to SUE framework, considering sensitivity of the results and toll rules to variations in the benefits of variety. The DUE model of Verhoef et al. (1996) assumes linear inverse demand:

$$D(N) = \delta_1 - \delta_2 N, \tag{36}$$

Table 1: Assumptions for calibrated paramaters in the deterministic user equilibrium.

δ_1	50
δ_2	0.01
κ_T	20
κ_U	20
β_T	0.02
β_U	0.02
α_1	0.8
α_2	1.3

and linear congestion costs for route r defined as:

$$c_r(N_r) = \kappa_r + \beta_r N_r, \tag{37}$$

where the base case assumed parameter values are summarized in Table 1. This implies that both routes are assumed to be identical in the base case resulting in non-intervention equilibrium route demands of $N_T = N_U = 750$. Substituting these values in equation (37) gives equilibrium average costs of $20 + 0.02 \times 750 = 35$ and marginal social costs of $20 + 0.04 \times 750 = 50$. Applying optimal first-best tolling results in average costs of 30 and marginal social costs of 40, whereas the toll is given by $40 - 30 = 10$. The socially optimal number of travellers is given by 500 for both routes.

In what follows we consider the various toll rules and welfare implications for homogeneous preferences and for heterogeneity in values of time between groups (α_k), and heterogeneity in preferences between groups (θ_k) for the symmetric case. In order to make a comparison between such cases, we calibrate the initial link flows at the non-intervention equilibrium to be equal to the UE non-intervention flows, and adjust the inverse demand function(s) accordingly. We also impose a constraint on the flow weighted average VOTs so that this is equal to the value used in the homogeneous case. This calibration of the model to observed flows and average VOTs not only ensures that the flows are consistent at the non-intervention case, but also that the initial aggregate welfare levels are maintained across models. The symmetric examples may hide some impacts, in particular the change in route flows when benefits of variety are included (even in the non-intervention case). For this reason we also develop an asymmetric example (again based on Verhoef et al. (1996)), where we adjust the systematic route utility with the introduction of a specific constant on one route, to maintain the equilibrium route flows. This asymmetric example is used to illustrate the fact that benefits of variety affect equilibrium tolls and flows even in the first-best homogeneous case. Calibration of this asymmetric case is directed to Appendix D.

When calibrating the SUE model with heterogeneous preferences, we purposely calibrate the demand function and initial group-specific demands such that the DUE non-intervention route flows (and demands) are retrieved as we move between cases. As we are dealing with linear inverse demand and cost functions, these adjustments are additive in nature. Equilibrium arises where marginal benefits are equal to the generalized price, and therefore we have to shift the inverse demand curve by an amount equal to the difference between the DUE average costs (35 at the no toll equilibrium in our example) and the no toll stochastic average costs. Hence the new inverse demand function can be written as:

$$D(N) = \delta_1 + \nu - \delta_2 N, \tag{38}$$

where the shift term ν is equal to $\frac{1}{\theta} \ln \left[\frac{N_T^0}{N^0} \right] < 0$ in order to maintain the same flows in unregulated equilibrium. Because the benefits of variety decrease expected costs, the inverse demand curve needs to be shifted downwards to maintain the same equilibrium non-intervention flows.

When we move to the case where groups are characterised by different values of time, we impose the condition that the flow-weighted average VOT is equal to 1, which is the assumed value with homogeneous travellers, reflecting that in the base calibration no explicit distinction between a travel time function and the average user cost function is made. In addition we maintain the initial demand, so that we have conditions as follows:

$$\sum_{k=1}^{K} \alpha_k \frac{N_k^0}{N^0} = 1; N^0 = \sum_{k=1}^{K} N_k^0, \tag{39}$$

where superscript N_k^0 refers to the non-intervention values of total group-specific demand and N^0 is total non-intervention demand. For the case with two groups we have:

$$N_1^0 = N^0 \frac{1 - \alpha_2}{\alpha_1 - \alpha_2}; N_2^0 = N^0 - N_1^0. \tag{40}$$

For the DUE case with two groups, we have to adjust the group-specific demand functions to account for the change in VOTs in order to maintain the initial systematic average costs as for the homogeneous case. This is achieved by adjusting the intercept and slope of the group-specific inverse demand curves as follows:

$$D_k(N_k) = \delta_1 \alpha_k + \nu_k - (\delta_1 - c_{due}^0) \alpha_k \frac{N_k}{N_k^0}, \tag{41}$$

where c_{due}^0 is the deterministic user equilibrium costs, and ν_k the negative correction term for the benefits of variety arising in SUE.

This procedure is best demonstrated by an example. Let the group-specific values of time be $\alpha_1 = 0.8$ and $\alpha_2 = 1.3$ (see Table 1). This gives initial flows of $N_1^0 = 1500 \frac{1-1.3}{0.8-1.3} = 900$ and $N_2^0 = 1500 - 900$ and intercepts of $\delta_1 \alpha_1 = 50 \times 0.8 = 40$ and $\delta_1 \alpha_2 = 50 \times 1.3 = 65$. Assuming $\nu_1 = \nu_2 = 0$ for the deterministic model, the slopes of the inverse demand curves for both groups are given by $-\frac{\alpha_1(\delta_1 - c_{due}^0)}{N_1^0} = -\frac{15 \times 0.8}{900} = -\frac{1}{75}$, and $-\frac{\alpha_2(\delta_1 - c_{due}^0)}{N_2^0} = -\frac{15 \times 1.3}{600} = -\frac{13}{400}$ respectively. Substituting these values back in the inverse demand function gives equilibrium non-intervention average costs of $40 - \frac{1}{75} \times 900 = 28$ and $65 - \frac{13}{400} \times 600 = 45\frac{1}{2}$. Note that their ratio corresponds to $\frac{1.3}{0.8}$, consistent with the equilibrium travel time being equal for both groups.

For the no toll DUE equilibrium, group-specific and total welfare can be calculated using these equilibrium average user costs. Because of linear inverse demand, welfare is given by the triangular area above the average cost curve. For group k the welfare is denoted by $W_{k,0}$. The group-specific welfare is given by $W_{1,0} = \frac{1}{2} \times (\delta_1 \alpha_1 - 28) \times N_1^0 = \frac{1}{2} \times 12 \times 900 = 5400$ and $W_{2,0} = \frac{1}{2} \times (\delta_1 \alpha_2 - 45\frac{1}{2}) \times N_2^0 = \frac{1}{2} \times 19\frac{1}{2} \times 600 = 5850$. Total initial welfare is then given by $\hat{W}_0 = W_{1,0} + W_{2,0} = 5400 + 5850 = 11250$, which is equal to the welfare of the homogeneous DUE case of Verhoef et al. (1996). This procedure for DUE can be extended to SUE by shifting the group-specific demand curves with the group-specific correction terms ν_1 and ν_2 in such a way that non-intervention average user costs and welfare levels are maintained.

5.2. First-best congestion pricing with homogeneous travellers and asymmetric route costs

We start with first-best tolling in the homogeneous VOT SUE model with symmetric route costs. When route costs are symmetric, the first-best tolls from equation (14) were found to give the same optimal tolls and flows as for the DUE case: tolls of 10 and flows of 500 on each link for all chosen values of θ. However, this is a special case, because it is with asymmetric route costs that θ has an effect on the toll via its impact on the equilibrium numbers of travellers on both routes, which in turn directly enters the first-best toll rules. To illustrate this, assume that the routes have different free-flow travel times κ_r, with $\kappa_T = 20$ and $\kappa_U = 10$. This changes the non-intervention flows on both routes and from Verhoef et al. (1996) these are $N_T^0 = 625$ and $N_U^0 = 1125$ for the DUE case. If we now introduce a preference for variety in the SUE case then these non-intervention flows would be different. As discussed in the previous section, we seek to maintain the observed route flows in the non-intervention case. Therefore we introduce a route specific constant for route U, which represents a route-specific preference not related to travel time and toll. In Appendix D we show that the required calibrated constant for any chosen value of θ amounts to:

$$ASC_U = \frac{1}{\theta} \ln \left[\frac{N_T^0}{N_U^0} \right], \tag{42}$$

where the flows are from the DUE non-intervention case. This results in a negative constant being added to the shorter route U, which attracts more users to compensate for the benefits of variety term.

Table (2) shows the main variables in the optimum for different values of θ, where the DUE case $\theta = \infty$ results in the equilibrium solution of $458\frac{1}{3}$ and $708\frac{1}{3}$ of Verhoef et al. (1996).

As opposed to the symmetric case, θ changes equilibrium route flows, which in turn impact the optimal toll levels. The optimal flows on the shorter route U, increase with a decreasing θ as the alternative specific constant ASC_U increases with decreasing θ. The flows on the longer route decrease when θ decreases. The tolls follow the flows as implied by the marginal first-best toll rules of (14). Compared to DUE, the overall demand and the corresponding welfare slightly reduce as the benefits of variety become more important (i.e. lower θ).

Table 2: Tolls, flows and welfare for FB tolling with asymmetric route costs.

θ	f_T	f_U	N_T	N_U	N	$\hat{W}_{FB}$
∞	9.17	14.17	458.33	708.33	1166.7	21042
10	9.16	14.17	458.11	708.49	1166.6	21041
1	9.12	14.20	456.21	709.85	1166.1	21039
0.5	9.09	14.22	454.26	711.23	1165.5	21036
0.1	8.86	14.38	443.11	719.03	1162.1	21020

5.3. First-best congestion tolling with observed heterogeneity in preferences

For first-best tolling with heterogeneous values of time and stochastic route choice we use group-specific values of time of $\alpha_1 = 0.8$ and $\alpha_2 = 1.3$. As described in section 5.1, we maintain the average VOT and welfare at the no toll equilibrium using initial flows of 900 and 600 respectively. It was confirmed numerically that the first-best tolls from equation (16) were optimal, and the resulting tolls, flows and welfare are shown in Table 3. When benefits of variety are important (low values of θ), there was only one solution with common first-best tolls which are higher than in the homogeneous case. As with the homogeneous case, the first-best toll solutions are independent of θ, due to the symmetry in average route costs for low values of θ. In DUE ($\theta \to \infty$), almost all the low VOTs group were priced off route T, with the remaining low VOT travellers using the untolled route.[7] Around 441 travellers of the higher VOT group use the tolled route and another 53.3 using the untolled route in the deterministic equilibrium. The total welfare is larger than for the homogeneous case, despite the total demand being only 974.3 users. This is due to the new average VOTs being 1.06 at the first-best equilibrium, because more high VOTs users enter the road.

Table 3: Tolls, flows and welfare for FB tolling with heterogeneous value of time

θ	f_{T1}	f_{T2}	f_{U1}	f_{U2}	N_{T1}	N_{T2}	N_{U1}	N_{U2}	$\hat{W}_{FB1}$	$\hat{W}_{FB2}$	$\hat{W}_{FB}$
∞	11.5	11.5	9.1	9.1	0	441.0	480	53.3	5888.0	9512.0	15400.0
10	11.4	11.4	9.1	9.1	0.3	437.5	471.8	60.4	5791.9	9558.4	15350.3
1	10.2	10.2	10.2	10.2	229.4	251.0	229.4	251.0	6081.7	9212.5	15294.1
0.5	10.2	10.2	10.2	10.2	229.4	251.0	229.4	251.0	6081.7	9212.5	15294.1

Note: for $\theta < 0.5$ the tolls, flows and welfare levels are exactly the same because we deal with symmetric route costs. Initial welfare for group 1 is given by 5400 and by 5850 for group 2.

For different values of θ two types of solutions arise with group flows tending towards a differentiated toll equilibrium with a high number of high VOTs users on the link with

[7] For high values of θ we obtain several solutions that satisfy the first-best toll expressions of equation (16) with heterogeneity. Due to symmetry we leave out two solutions because it is always possible to swap the route flows. Furthermore, there may be solutions that satisfy the toll expression but that are local minima. We therefore used different starting values and checked the eigenvalues of the Hessian matrix to be sure that the solution found is a maximum.

a high toll, and a high number of low VOTs users at the other route. This separation result occurs also in the heterogeneous DUE case of Arnott et al. (1992) and Verhoef and Small (2004), but eventually disappears in the SUE case when θ becomes sufficiently low. Route preferences of individuals then become so stochastic that toll differentiation is not beneficial in welfare terms. The toll differentiated equilibrium then dissipates due to the lower sensitivity to deterministic costs, and the toll differentiated solution is "smoothed" out, by randomness in route choice. Toll differentiated equilibria in our model become more likely for two reasons. First, these equilibria become more likely when route choice is more deterministic, so for higher values of θ. Second, when values of times are more heterogeneous, a toll differentiated equilibrium is more likely to occur because it is more beneficial to offer differentiated roads (see Small and Yan (2001) and Verhoef and Small (2004)). For our model this implies that when we increase the difference between α_1 and α_2, while keeping the average VOTs constant, a toll differentiated equilibrium will occur for lower values of θ (numerical results available upon request).

Including benefits of variety also changes the distributional impacts of congestion tolls. When route choice is almost deterministic ($\theta = 10$), the low VOT group has a welfare gain of 391.9, whereas the high VOT group benefits more with an increase of 3708.4 in welfare. This increase is lower for the low VOT group compared to the DUE. For values of $\theta < 10$ we observe a higher welfare increase for the low VOT group compared to the DUE case.

5.4. Second-best tolling, homogeneous values of time

When first-best tolling is not feasible, second-best tolling with a toll on route T might be a realistic and viable alternative. The numerical results in this section confirm the optimal toll rule of equation (25). Figure (1) shows how the welfare improvement varies with the second-best toll on route T for different values of θ. Quite intuitively, as θ increases, the solution of the second-best toll tends towards the UE solution of 5.45 of Verhoef et al. (1996). The general tendency in Figure (1) is that the optimal second-best toll increases when route preferences become more stochastic. The reason is that travellers are less responsive to the deterministic part of utility, and therefore the behavioural response to the toll to route T is less strong. This allows the regulator to more fully internalize the marginal external costs on route T, without spillovers upon route U mitigating the gains, and therefore SB tolls can be higher when the benefits of variety increases. Randomness in utility thus mitigates the central inefficiency under second-best tolling. Table (4) shows the optimal second-best tolls, route flows, and relative efficiencies $\omega = \frac{\hat{W}_{SB} - \hat{W}_0}{\hat{W}_{FB} - \hat{W}_0}$. The latter is defined as the welfare gain due to second-best regulation divided by the welfare gain due to first-best regulation, where non-intervention is the benchmark (see Verhoef et al. (1995)).

As expected, the optimal toll with SUE increases when θ decreases, because road users become less sensitive to the deterministic part of average costs. Because equilibrium expected generalised costs increase with θ, overall demand decreases as well. The relative efficiency increases with decreasing θ, as the induced welfare losses on route U become smaller. This implies that the welfare losses due to second-best congestion pricing are lower when route choice is governed by random utility maximization.

Figure 1: Welfare gains $\hat{W}_{SB} - \hat{W}_0$ against toll level f_T for second-best tolling with homogeneous values of times for varying values of θ.

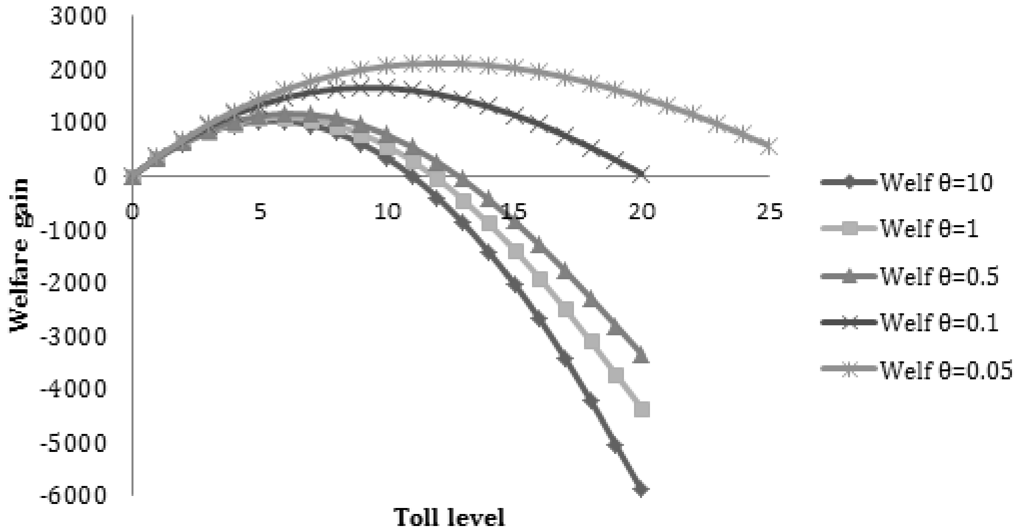

Table 4: Tolls, route flows and welfare for second-best tolling with symmetric route costs and homogeneous values of times.

θ	f_T	N_T	N_U	N	$\hat{W}_{SB} - \hat{W}_0$	ω
∞	5.45	545.00	818.18	1363.60	1022.70	0.27
10	5.50	544.95	817.74	1362.70	1029.90	0.27
1	5.87	540.63	813.67	1354.30	1093.90	0.29
0.5	6.28	536.18	808.96	1345.14	1163.30	0.31
0.1	9.26	510.43	768.57	1279.00	1653.70	0.44
0.05	12.13	493.56	721.03	1214.60	2113.20	0.56

5.5. Second-best tolling, heterogeneous benefits of variety

Next, we allow for heterogeneity in the scale of utility, and thus in the importance of unobserved preferences, between two groups of equal size and with their VOTs set to 1. We will compare the results with the homogeneous second-best toll case of the previous section. The example follows the symmetric case, where θ_2 is varied for group 2 holding θ_1 constant at 10. This allows us to study the effect of heterogeneous benefits of variety. Demand is calibrated as discussed in section (5.1). As we deal with symmetric route costs, the initial group flows are split equally between the links in the no toll case. Table (5) shows the results for the second-best group-specific tolls from equation (32), which were confirmed numerically to give the optimal tolls. The second row of Table (5) shows the result for homogeneous benefits of variety and has the same toll as the toll in the second row of Table (4). Table (5) shows the total welfare so that we can examine the differences between groups. The base welfare is 11250 so the total welfare gain corresponds to the reported value in Table (4) second line. A decrease in θ_2 results in a decrease of the optimal toll for group 1 and an increase of the optimal toll for group 2. Since group 2 has higher benefits of variety, this group is less responsive to the toll and a higher toll can be charged without causing serious spillovers, and more of the congestion externalities of route T can be internalized.

Table 5: Tolls, flows and welfare for SB tolling with heterogeneous returns to variety

θ_1	θ_2	f_{T1}	f_{T2}	N_{T1}	N_{T2}	N_{U1}	N_{U2}	N_1	N_2	$\hat{W}_{SB1}$	$\hat{W}_{SB2}$	$\hat{W}_{SB}$
10	∞	5.51	5.45	238.9	306.4	442.0	375.8	680.9	682.2	5935.0	6323.8	12277
10	10	5.50	5.50	272.5	272.5	408.9	408.9	681.4	681.4	6140.0	6140.0	12280
10	1	5.44	5.98	300.5	243.4	384.5	429.9	685.0	673.3	6325.3	5989.4	12315
10	0.5	5.37	6.51	304.1	238.9	384.8	425.8	688.9	664.7	6380.1	5972.3	12352
10	0.1	4.94	10.07	311.4	226.0	404.8	378.4	716.2	604.4	6668.4	5928.5	12597

Note: Initial welfare levels for group 1 and 2 are given by $\frac{11250}{2} = 5625$.

The tolls for group 2 are consistently higher than those for the same value of θ in the homogeneous case in Table (4), because the said mechanism prevails whenever $\theta_2 < 10$. Consistent with the toll levels, the equilibrium flows on the tolled link for group 1(2) increases (decreases) as θ_2 decreases for group 2. The group-specific welfare levels show that group 1 benefits from the decrease of θ_2. The result that the toll is higher for the second group as θ_2 decreases can be inferred from the toll rule of equation (32), where the second term represents the group-specific route substitution and demand effects. This term increases with decreasing own values of θ, so that the toll increases whereas the term decreases for decreasing θ of other groups, so that then the toll decreases.

Heterogeneity in the benefits of variety gives rises to another type of distribution effects when group specific second-best tolls are applied: the group with highest benefits of variety pays the highest toll and also has the lowest welfare gains. Both results stem from the untolled route being a less attractive alternative. The group with more deterministic preferences benefits from this and pays a lower toll and has higher welfare gains when the benefits of variety of the other group increase.

5.6. Second-best congestion tolling, heterogeneous values of time, group-specific tolls

This section presents the results for group-specific second-best tolls with heterogeneous values of times of $\alpha_1 = 0.8$ and $\alpha_2 = 1.3$. Table (6) presents the numerical results for different values of the scale parameter θ which is assumed to be equal across groups for this example. The SB tolls for both groups first decrease in θ and then increases in θ for lower values of θ. A lower θ means that more travellers with a low VOT and fewer travellers with a high VOT will use the tolled route. This leads to a downward adjustment of the first direct term in equation (32), which captures the marginal external costs of route T. But a further decrease in θ also means that spillovers become less and less important, and that means that the second term in equation (32) decreases. This effect raises the value of the second-best toll. The U-shaped pattern in Table (6) is the combined result of these two opposing forces.

Table 6: Tolls, route flows and welfare for differentiated second-best tolls with heterogeneous values of time

θ	f_{T1}	f_{T2}	N_{T1}	N_{T2}	N_{U1}	N_{U2}	N_1	N_2	$\hat{W}_{SB1}$	$\hat{W}_{SB2}$	$\hat{W}_{SB}$
∞	8.55	8.55	0.00	503.4	801.4	30.8	801.4	534.3	4281.3	8941.3	13223
10	8.28	8.55	0.00	496.2	793.1	41.7	793.1	537.9	4193.2	8947.0	13140
1	7.37	7.72	89.1	404.2	675.2	152.3	764.3	556.5	4550.7	8151.3	12702
0.5	7.31	7.25	175.5	331.2	583.6	230.1	759.1	561.3	5125.3	7521.3	12647
0.1	11.0	8.45	230.1	263.0	450.2	307.6	680.3	570.6	5604.9	7512.7	13118
0.05	14.7	10.4	218.8	254.8	381.2	319.3	600.0	574.0	5624.6	8008.2	13633

Note: Initial welfare for group 1 is given by 5400 and by 5850 for group 2.

Including benefits of variety gives interesting insights on the distributional impacts of second-best congestion tolling. When benefits of variety are low, the high VOT group will gain whereas the low VOT group will lose from second-best tolling. However, when benefits of variety increase the low VOT group might benefit from congestion tolling as well. Therefore there are cases where both groups benefit from second-best congestion which is the result of having two dimensions of heterogeneity in the model: tolling will impact both travel time costs as well as benefits of variety. When benefits of variety are high a more equal distribution of travellers over the routes is more beneficial. In the presence of high benefits of variety, the low VOT group is less responsive to the toll and a higher toll is needed to reduce total demand of this group. For practical implementation the case where both groups benefit is likely not political acceptable because the SB toll of the low VOT group is higher than the SB toll for the high VOT group.

5.7. Second-best congestion tolling, heterogeneous values of time, common tolls

It may well be that the regulator is not able to observe and distinguish the VOTs for different groups and that only undifferentiated "common" tolls can be applied. Table (7) shows that the numerically determined common second-best tolls are between the group-differentiated second-best tolls of Table (6). Because tolls cannot be differentiated between

Table 7: Tolls, flows and welfare for common second-best tolls with heterogeneous values of time

θ	f_T	N_{T1}	N_{T2}	N_{U1}	N_{U2}	N_1	N_2	$\hat{W}_{SB1}$	$\hat{W}_{SB2}$	$\hat{W}_{SB}$
10	8.55	0.0	496.2	793.1	41.7	793.1	537.9	4193.2	8947.0	13140
1	7.60	73.2	420.0	688.9	139.1	493.2	828.0	4427.3	8271.3	12699
0.5	7.28	177.9	328.9	581.8	231.9	506.8	813.7	5142.1	7504.4	12646
0.1	9.81	258.1	236.0	448.5	313.7	706.6	549.7	5860.9	7225.1	13086
0.05	12.77	251.9	224.2	395.8	313.6	647.7	537.9	6013.7	7565.8	13580

Note: Initial welfare for group 1 is given by 5400 and by 5850 for group 2.

groups, overall welfare levels can never be higher than in the previous section. When we compare differentiated and common second-best tolls for higher values of θ, the high VOT users benefit further from a common toll for higher values of θ, but benefit less for lower values of θ. Compared to group-specific SB tolls, common tolls therefore appear to benefit the high VOT users when route choice is more deterministic and lead to higher welfare gains compared for the low VOT users for high benefits of variety. When comparing the results of Table (6) with Table (7) we find that for $\theta = 0.5$, the second-best tolls of both groups and the corresponding welfare gains are almost equal, meaning that for this case differentiation of tolls between groups is hardly beneficial. The welfare benefits of toll differentiation are therefore influenced by the randomness of route preferences, and are highest for very low and very high values of θ. This nuances earlier findings on the welfare benefits of toll differentiation in deterministic models (see Verhoef and Small (2004); Small and Yan (2001)), and calls for empirical investigation of scale heterogeneity for practical tolling applications.

The distributional impacts of SB common tolls are comparable to the case with differentiated SB tolls. For high benefits of variety both groups might gain from congestion tolling whereas for more deterministic route choice only the high VOT group gains. For all cases presented the high VOT group always benefits more in absolute terms.[8]

6. Extension to larger networks

Having now considered various variants of the two-route problem, a natural follow-up question is to what extent our results may be expected to carry over to more general net-

[8]For high values of θ the solution approaches the DUE case reported in the first line of Table 5. It is then computationally difficult to find a solution as the route flow of group 1 on the tolled route becomes very small (see Clark et al. (2009) for a discussion on these issues for DUE network models with heterogeneity). This shows another advantage of reformulating the constrained optimization problem into an unconstrained optimization problem as done for the other cases in the paper.

works. A full analysis of this question seems beyond the scope of this paper, but quite some intuition can be gained by considering this issue in the context of the generalization of Verhoef et al. (1996) as it was presented in Verhoef (2002a,b). This work generalized the deterministic two-route problem to the problem of finding the second-best optimal toll formulae for a deterministic general network, of undetermined size and shape, on which an arbitrary sub-set of links can be tolled. The mathematical formulation of the problem thus allows for a set of OD pairs, indexed i; each of which has its own inverse demand function D_i and a set of possible routes or paths p where an indicator δ_{ip} indicates that path p serves OD-pair i; and a set of links indexed j, each with their own average cost function c_j, where an indicator δ_{jp} indicates that link j belongs to path p. Like in the two-route problem considered above, the associated Lagrangian consists of an objective that adds up user benefits over all OD-pairs and subtracts user costs added up over all links, and constraints that equate marginal benefit to the generalized price for every used route. Without specifying the network's graph, no closed-form solutions exist for the associated route-specific Lagrangian multipliers λ_p, but the second-best optimality conditions for these multipliers as well as the available second-best tolls f_l, all in function of other route-specific multipliers $\lambda_{q \neq p}$, are nevertheless instructive and will also be helpful in the current context. Not surprisingly, this implicit solution has the solution to the deterministic two-route problem, and the associated Lagrangian multipliers and second-best toll, as a special case.

To trace the consequences of extending the basic model of Section 3.1 to cover a full network, first observe that the objective function in Equations 6 and 7 would be extended by having summations of the conventional Marshallian benefit terms and the entropy terms over OD-pairs i, where for each such term the relevant N's are those route flows that pertain to the OD-pair under consideration. Next, user cost terms would be included for all links j, where for each link the relevant N's include all those route flows that use the link under consideration. Now to the extent that idiosyncratic preferences can be assumed to pertain to routes, or can be written as such after summing link-specific idiosyncratic terms over the route's links, this means that in the second-best optimality conditions for the resulting Lagrangian (e.g. equation (6-9) in Verhoef (2002a)), and therefore also in the associated expressions for path-specific multipliers λ_p and second-best link-specific tolls f_l (e.g. equations (10) and (11) in Verhoef (2002a)), we can expect marginal entropy (or θ-related) terms to show up in perfect companion with OD-specific marginal willingness-to-pay-related (or D-related) terms. These marginal entropy terms will take on similar forms as in Equations (9), (12) and (13) in the current paper. In contrast, the structure of interactions between link-specific costs will remain unaltered compared to how these work out on the deterministic generalized networks of Verhoef (2002a,b).

Given the complexity of the general second-best tax expression obtained in Verhoef (2002a,b), where direct and indirect demand and cost interactions will eventually occur between any pair of links or OD-pairs, it is impossible to describe exactly how the consideration of idiosyncratic preferences would alter the insights obtained for deterministic networks. But given the limiting impact that idiosyncratic preferences have on route diversion in response to second-best tolls, and given the maintained purely additive (over relevant serial links) generalized cost expressions for systematic generalized costs and tolls at the route level, one would expect in larger networks that the consideration of idiosyncratic preferences leads in particular to a relatively strong upward adjustment of second-best tolls where these were lowered under deterministic preferences to prevent excessive induced congestion on unpriced parallel routes. In contrast, one would expect relatively modest upward adjustments for second-best link-tolls in corridors that offer no (congested) alternatives. That is: interactions between links that are largely substitutes, or that are parts of routes or route segments that are largely substitutes, can be expected to justify smaller downward toll adjustments with idiosyncratic preferences than with deterministic preferences. Interactions between links that are largely (serial) complements will justify more modest adaptations because of idiosyncratic preferences – ignoring now, of course, toll adjustments on any such pair of links that is part of a corridor that would face reduced spill-overs from parallel competition after introducing idiosyncratic preferences. And finally, since idiosyncratic preferences primarily limit substitution under second-best pricing which in itself usually justifies downward toll adjustments, one might expect upward toll adjustments to dominate when introducing idiosyncratic preferences. But exactly because of network interactions, we hypothesize that incidental downward toll adjustments cannot be excluded. An example could be the case where under deterministic preferences an upstream toll was relatively high because a downstream unpriced-substitute problem led to a low downstream second-best toll. In such cases, the downstream toll increase that would result from considering idiosyncratic preferences might well justify an upstream toll reduction, especially if that upstream link is also used by users who do not use the described downstream trip segment. Obviously, hypotheses like these justify further study for more realistic general networks.

It seems impossible to give any more specific predictions without narrowing down the size and shape of the network and considering alternative archetype configurations, which seems hard to fit within the scope of this paper. We therefore leave further investigation of this issue for future work.

7. Conclusion

This paper presented new analytical results for optimal first-best and second-best congestion prices in the presence of observed and unobserved preference heterogeneity. It revisited the classical two route problem of Verhoef et al. (1996) and extends it to include two dimensions of heterogeneity: heterogeneous values of travel times and heterogeneity in the degree of unobserved route preferences. Our analytical approach incorporates travellers' benefits of variety in a tractable way in the welfare function and provides new insights on potential distributional impacts of congestion pricing and the value of the ability to differentiate taxes between alternative routes or modes. The analytical approach that we used can be useful for other discrete choice applications where externalities are present. Extensions to an arbitrary number of alternatives are presented in the paper. It is a first step towards the analysis of taxation of externalities in networks with general stochastic and deterministic preference heterogeneity.

We show that when values of travel time savings are homogeneous, welfare losses due to second-best pricing are lower when benefits of variety are present than for deterministic route choice. When values of travel times are heterogeneous, the picture is less clear cut because the benefits of toll differentiation between groups first decrease, and then increase when travellers value variety more. In line with this, we find that there are cases where the non-differentiated second-best congestion tolls are very close to the group-specific (differentiated) tolls. For these cases the welfare loss due to the inability to differentiate congestion taxes between groups is negligible, which makes implementation of these taxes much easier because users can be treated as anonymous.

One assumption we make in this paper is that we have an equilibrium in expected utility implying that travellers only learn the realisation of the stochastic part of utility after making their choice. Therefore the best they can do to optimise their decisions is to optimize their expected utility. When travellers know the stochastic part of utility before making the choice, the random part of utility will enter the constraints in Equation (7) instead of the entropy terms.

For future research it is interesting to investigate the size of the benefits of variety empirically, and to investigate whether our qualitative results also hold for realistic large scale networks as discussed in more detail in Section 6. This will allow us to make more precise quantitative predictions about the impact of stochastic user equilibrium on tolling policy recommendations.

Bibliography

Anderson, S., Palma, A. D., and Thisse, J. (1992). *Discrete choice theory of product differentiation.* MIT Press, Cambridge (MA).

Anderson, S. P., Palma, A. D., and Thisse, J. F. (1988). A representative consumer theory of the logit model. *International Economic Review*, 29:461–466.

Arnott, R., de Palma, A., and Lindsey, R. (1992). Route choice with heterogeneous drivers and group-specific congestion costs. *Regional Science and Urban Economics*, 22(1):71–102.

Arnott, R., de Palma, A., and Lindsey, R. (1994). The Welfare Effects of Congestion Tolls with Heterogeneous Commuters. *Journal of Transport Economics and Policy*, 28:139–161.

Arnott, R. and Kraus, M. (1998). When are anonymous congestion charges consistent with marginal cost pricing? *Journal of Public Economics*, 67(1):45–64.

Bayer, P. and Timmins, C. (2007). Estimating equilibrium models of sorting across locations. *Economic Journal*, 117(518):353–374.

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

Daganzo, C. F. and Sheffi, Y. (1977). On Stochastic Models of Traffic Assignment. *Transportation Science*, 11(3):253–274.

de Jong, G., Daly, A., Pieters, M., and van der Hoorn, T. (2007). The logsum as an evaluation measure: Review of the literature and new results. *Transportation Research Part A: Policy and Practice*, 41:874–889.

de Palma, A., Kilani, M., and Lindsey, R. (2008). The merits of separating cars and trucks. *Journal of Urban Economics*, 64(2):340–361.

Erlander, S. (1977). Accessibility, Entropy and the Disruption and Assignment of Traffic. *Transportation Research*, 11(3):149–153.

Fisk, C. (1980). Some developments in equilibrium traffic assignment. *Transportation Research Part B: Methodological*, 14(3):243–255.

Huang, H. J. and Li, Z. C. (2007). A multiclass, multicriteria logit-based traffic equilibrium assignment model under ATIS. *European Journal of Operational Research*, 176(3):1464–1477.

Jiang, L., Mahmassani, H., and Zhang, K. (2011). Congestion pricing, heterogeneous users, and travel time reliability: multicriterion dynamic user equilibrium model and efficient implementation for large-scale networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2254(1):58–67.

Lévy-Lambert, H. (1968). Tarification des Services à Qualité Variable–Application aux Péages de Circulation. *Econometrica: Journal of the Econometric Society*, 36:564–574.

Lu, C., Zhou, X., and Mahmassani, H. (2006). Variable toll pricing and heterogeneous users: Model and solution algorithm for bicriterion dynamic traffic assignment problem. *Transportation Research Record: Journal of the Transportation Research Board*, 1964:19–26.

Luce, D. and Suppes, P. (1965). Preference, utility and subjective probability. In Luce, R. Galanter, E., Bush, R., editor, *Handbook of Mathematical Psychology*, pages 249–410. John Wiley and Sons, New York.

Maher, M., Stewart, K., and Rosa, A. (2005). Stochastic social optimum traffic assignment. *Transportation Research Part B: Methodological*, 39:753–767.

Mahmassani, H., Zhou, X., and Lu, C.-C. (2005). Toll Pricing and Heterogeneous Users: Approximation Algorithms for Finding Bicriterion Time-Dependent Efficient Paths in Large-Scale Traffic Networks. *Transportation Research Record*, 1923:28–36.

Marchand, M. (1968). A Note on Optimal Tolls in an Imperfect Environment. *Econometrica*, 36:575–581.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in econometrics*, pages 105–142. Academic Press, New York.

Miyagi, T. (1986). On the Formulation of a Stochastic User Equilibrium Model Consistent with the Random Utility Theory. A Conjugate Dual Approach. *Research for Tomorrow's Transport Requirements. Proceedings of the Fourth World Conference on Transport Research*, 2:1619–1635.

Nash, C. (2003). Project UNITE: UNIfication of accounts and marginal costs for Transport Efficiency. Final

Technical Report, Fifth Framework Competitive And Sustainable Growth Programme, Commissioned by European Commission, DG TREN. (with contributions from partners). Technical report, ITS Leeds.

Parry, I. W. H. (2008). How should heavy-duty trucks be taxed? *Journal of Urban Economics*, 63(2):651–668.

Pigou (1920). *The economics of welfare*. MacMillan, London.

Small, K., Verhoef, E., and Lindsey, R. (2007). *The economics of urban transportation*. Routledge.

Small, K. A. (2012). Valuation of travel time. *Economics of Transportation*, 1:2–14.

Small, K. A. and Rosen, H. S. (1981). Applied Welfare Economics with Discrete Choice Models. *Econometrica*, 49:105–130.

Small, K. A. and Yan, J. (2001). The Value of Value Pricing of Roads: Second-Best Pricing and Product Differentiation. *Journal of Urban Economics*, 49:310–336.

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

United Nations (2014). World Urbanization Prospects, The 2014 revision. Technical Report. Technical report, Department of Economic and Social Affairs, Population Division.

van den Berg, V. and Verhoef, E. T. (2011a). Congestion tolling in the bottleneck model with heterogeneous values of time. *Transportation Research Part B: Methodological*, 45:60–78.

van den Berg, V. and Verhoef, E. T. (2011b). Winning or losing from dynamic bottleneck congestion pricing?. The distributional effects of road pricing with heterogeneity in values of time and schedule delay. *Journal of Public Economics*, 95:983–992.

van den Berg, V. and Verhoef, E. T. (2013). Congestion pricing in a road and rail network with heterogeneous values of time and schedule delay. *Transportmetrica A: Transport Science*, 10(5):377–400.

Verboven, F. (1996). The nested logit model and representative consumer theory. *Economics Letters*, 50:57–63.

Verhoef, E., Nijkamp, P., and Rietveld, P. (1995). Second-Best Regulation of Road Transport Externalities. *Journal of Transport Economics and Policy*, 29:147–167.

Verhoef, E., Nijkamp, P., and Rietveld, P. (1996). Second-best congestion pricing: The case of an untolled alternative. *Journal of Urban Economics*, 40(3):279–302.

Verhoef, E. T. (2002a). Second-best congestion pricing in general networks. Heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research Part B: Methodological*, 36:707–729.

Verhoef, E. T. (2002b). Second-best congestion pricing in general static transportation networks with elastic demands. *Regional Science and Urban Economics*, 32:281–310.

Verhoef, E. T. and Small, K. A. (2004). Product Differentiation on Roads: Constrained Congestion Pricing with Heterogeneous Users. *Journal of Transport Economics and Policy*, 38:127–156.

Walters, A. A. (1961). The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica*, 29(4):676–699.

Williams, H. C. W. L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9(3):285–344.

Yang, H. (1999). System optimum, stochastic user equilibrium, and optimal link tolls. *Transportation Science*, 33(4):354–360.

Yang, H. and Huang, H. J. H. J. (2004). The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. *Transportation Research Part B: Methodological*, 38:1–15.

Zhang, L., Levinson, D. M., and Zhu, S. (2008). Agent-based model of price competition, capacity choice, and product differentiation on congested networks. *Journal of Transport Economics and Policy*, 42:435–461.

Appendix A. Derivation of first-best tolls with heterogeneous preferences

Define $\bar{N_j}^{\alpha} = \sum_{k=1}^{K} \alpha_k N_{jk}$ as the preference weighted average number of travellers at route j. Then the Lagrangian is given by:

$$\begin{aligned}\mathcal{L} = &\sum_{k=1}^{K} \int_0^{N_k} D_k(n_k)dn_k - \sum_{j=1}^{J} \bar{N_j}^{\alpha} c_j(N_j) - \sum_{k=1}^{K} \frac{1}{\theta_k} \sum_{j=1}^{J} \left(N_{jk} \ln \left[\frac{N_{jk}}{N_j} \right] \right) \\ &+ \sum_{k=1}^{K} \sum_{j=1}^{J} \lambda_{jk} \left(D_k(N_k) - f_{jk} - \alpha_k c_j(N_j) - \frac{1}{\theta_k} \ln \left[\frac{N_{jk}}{N_k} \right] \right)\end{aligned} \tag{A.1}$$

The first-order conditions are given by:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial N_{il}} &= D_l(N_l) - \bar{N_i}^{\alpha} c_i'(N_i) - \alpha_l c_i(N_i) - \frac{1}{\theta_l} \ln \left[\frac{N_{il}}{N_l} \right] \\ &- \sum_{k=1}^{K} \sum_{j=1}^{J} \lambda_{jk} \alpha_k c_i'(N_i) = 0, \forall i = 1...J, \forall l = 1...K.\end{aligned} \tag{A.2}$$

$$\frac{\partial \mathcal{L}}{\partial f_{il}} = -\lambda_{il} = 0, \forall i = 1...J, \forall l = 1...K. \tag{A.3}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_{il}} = D_l(N_l) - f_{il} - \alpha_l c_i(N_i) - \frac{1}{\theta_l} \ln \left[\frac{N_{il}}{N_l} \right] = 0, \forall i = 1...J, \forall l = 1...K. \tag{A.4}$$

Equations (A.3) show that the Lagrangian multipliers are 0. Substituting equations (A.3) and (A.4) in equation (A.2) we obtain:

$$f_{il} = \bar{N_i}^{\alpha} c_i'(N_i), \forall i = 1..J, \forall l = 1...K. \tag{A.5}$$

Because $\bar{N_i}^{\alpha}$ is equal for all groups, the tolls on every route are equal for all groups. This is because every additional traveller raises congestion with the same amount.

Appendix B. Second-best pricing with homogeneous preferences

For second-best pricing with homogeneous preferences we define the set of routes as S, the subset of tolled routes as S_T and the subset of untolled routes as S_U. The Lagrangian is given by:

$$\begin{aligned}L = &\int_0^{N} D(n)dn - \sum_{j \in S} N_j c_j(N_j) - \frac{1}{\theta} \sum_{j \in S} N_j \ln[N_j] + \sum_{j \in S_T} \lambda_j \left(D(N) - \frac{1}{\theta} \ln \left[\frac{N_j}{N} \right] - f_j - c_j(N_j) \right) \\ &+ \sum_{j \in S_U} \lambda_j \left(D(N) - \frac{1}{\theta} \ln \left[\frac{N_j}{N} \right] - c_j(N_j) \right).\end{aligned} \tag{B.1}$$

For the tolled routes the first-order conditions are given by:

$$\begin{aligned}\frac{\partial L}{\partial N_i} &= D(N) - c_i(N_i) - N_i c_i'(N_i) - \frac{1}{\theta}\ln\left[\frac{N_i}{N}\right] + \lambda_i\left(D'(N) - \frac{1}{\theta}\frac{N-N_i}{NN_i} - c_i'(N_i)\right) \\ &+ \sum_{j\in S_T, j\neq i}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) + \sum_{j\in S_U}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = 0, \forall i \in S_T.\end{aligned} \tag{B.2}$$

For the untolled routes the first-order conditions are given by:

$$\begin{aligned}\frac{\partial L}{\partial N_i} &= D(N) - c_i(N_i) - N_i c_i'(N_i) - \frac{1}{\theta}\ln\left[\frac{N_i}{N}\right] + \lambda_i\left(D'(N) - \frac{1}{\theta}\frac{N-N_i}{NN_i} - c_i'(N_i)\right) \\ &+ \sum_{j\in S_U, j\neq i}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) + \sum_{j\in S_T}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = 0, \forall i \in S_U.\end{aligned} \tag{B.3}$$

$$\frac{\partial L}{\partial f_i} = -\lambda_i = 0, \forall i \in S_T \tag{B.4}$$

$$\frac{\partial L}{\partial \lambda_i} = D(N) - \frac{1}{\theta}\ln\left[\frac{N_i}{N}\right] - f_i - c_i(N_i) = 0, \forall i \in S_T \tag{B.5}$$

$$\frac{\partial L}{\partial \lambda_i} = D(N) - \frac{1}{\theta}\ln\left[\frac{N_i}{N}\right] - c_j(N_j) = 0, \forall i \in S_U \tag{B.6}$$

From B.4 we know that the multipliers for the tolled routes are 0. Together with B.5 and B.6 we can rewrite B.2 and B.3.

$$\begin{aligned}&\frac{\partial L}{\partial N_i} = f_i - N_i c_i'(N_i) + \sum_{j\in S_U}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = 0, \forall i \in S_T. \\ &\Leftrightarrow \\ &f_i = N_i c_i'(N_i) - \sum_{j\in S_U}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right), \forall i \in S_T.\end{aligned} \tag{B.7}$$

This shows that when we solve for the multipliers of the untolled routes we obtain a closed-form expression for the second-best toll.

$$\frac{\partial L}{\partial N_i} = -N_i c_i'(N_i) + \lambda_i\left(D'(N) - \frac{1}{\theta}\frac{N-N_i}{NN_i} - c_i'(N_i)\right) + \sum_{j\in S_U, j\neq i}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = 0, \forall i \in S_U. \tag{B.8}$$

Using $\frac{1}{\theta}\frac{N-N_i}{NN_i} = \frac{1}{N_i} - \frac{1}{N}$ gives:

$$\lambda_i\left(-\frac{1}{\theta}\frac{1}{N_i} - c_i'(N_i)\right) + \sum_{j\in S_U}\lambda_j\left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) = N_i c_i'(N_i), \forall i \in S_U. \tag{B.9}$$

Solving for λ_i gives:

$$\lambda_i = \frac{\sum_{j \in S_U} \lambda_j \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right)}{\left(\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)\right)} - \frac{N_i c_i'(N_i)}{\left(\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)\right)}, \forall i \in S_U. \tag{B.10}$$

Summing these multipliers over all untolled routes gives:

$$\begin{aligned}
\sum_{i \in S_U} \lambda_i &= \sum_{i \in S_U} \left(\frac{\sum_{j \in S_U} \lambda_j \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right)}{\left(\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)\right)} - \frac{N_i c_i'(N_i)}{\left(\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)\right)} \right), \forall i \in S_U. \\
&= \sum_{i \in S_U} \sum_{j \in S_U} \lambda_j \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) \left(\frac{1}{\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)} \right) - \sum_{i \in S_U} \frac{N_i c_i'(N_i)}{\frac{1}{\theta}\frac{1}{N_i} - c_i'(N_i)}, \forall i \in S_U. \\
&= \left(\sum_{j \in S_U} \lambda_j \right) \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) \sum_{i \in S_U} \left(\frac{1}{\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)} \right) - \sum_{i \in S_U} \frac{N_i c_i'(N_i)}{\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)}, \forall i \in S_U. \\
&\Leftrightarrow \\
&\left(\sum_{j \in S_U} \lambda_j \right) \left(1 - \sum_{i \in S_U} \frac{D'(N) + \frac{1}{\theta}\frac{1}{N}}{\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)} \right) = - \sum_{i \in S_U} \frac{N_i c_i'(N_i)}{\frac{1}{\theta}\frac{1}{N_i} + c_i'(N_i)}, \forall i \in S_U.
\end{aligned} \tag{B.11}$$

Solving for $\sum_{j \in S_U} \lambda_j$ gives:

$$\sum_{j \in S_U} \lambda_j = \frac{\sum_{j \in S_U} \frac{N_j c_j'(N_j)}{\frac{1}{\theta}\frac{1}{N_j} + c_j'(N_j)}}{\sum_{j \in S_U} \frac{D'(N) + \frac{1}{\theta}\frac{1}{N}}{\frac{1}{\theta}\frac{1}{N_j} + c_j'(N_j)} - 1} \tag{B.12}$$

Which is negative under the assumption that $D'(N) + \frac{1}{\theta}\frac{1}{N} < 0$ and $c_j'(N_j) > 0$. Substituting in B.7 gives:

$$f_i = N_i c_i'(N_i) - \left(D'(N) + \frac{1}{\theta}\frac{1}{N}\right) \frac{\sum_{j \in S_U} \frac{N_j c_j'(N_j)}{\frac{1}{\theta}\frac{1}{N_j} + c_j'(N_j)}}{\sum_{j \in S_U} \frac{D'(N) + \frac{1}{\theta}\frac{1}{N}}{\frac{1}{\theta}\frac{1}{N_j} + c_j'(N_j)} - 1}, \forall i \in S_T \tag{B.13}$$

Appendix C. Derivation of group-specific second-best tolls with heterogeneous preferences

Define the set of routes as S, the subset of tolled routes as S_T and the subset of untolled routes as S_U. The Lagrangian is given by:

$$\begin{aligned}\mathcal{L} = &\sum_{k=1}^{K}\int_0^{N_k} D_k(n_k)dn_k - \sum_{k=1}^{K}\sum_{j\in S}\alpha_k N_{jk}c_j(N_j) - \sum_{k=1}^{K}\sum_{j\in S}\frac{1}{\theta_k}N_{jk}\ln\left[\frac{N_{jk}}{N_k}\right]\\ &+\sum_{k=1}^{K}\sum_{j\in S_T}\lambda_{jk}\left(D_k(N_k) - f_{jk} - \alpha_k c_j(N_j) - \frac{1}{\theta_k}\ln\left[\frac{N_{jk}}{N_k}\right]\right)\\ &+\sum_{k=1}^{K}\sum_{j\in S_U}\lambda_{jk}\left(D_k(N_k) - \alpha_k c_j(N_j) - \frac{1}{\theta_k}\ln\left[\frac{N_{jk}}{N_k}\right]\right)\end{aligned} \tag{C.1}$$

For tolled routes we obtain the following first-order condition for group ℓ for route i:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial N_{i\ell}} = &D_\ell(N_\ell) - \sum_{k=1}^{K}\alpha_k N_{ik}c_i'(N_i) - \alpha_\ell c_j(N_j) - \frac{1}{\theta_\ell}\ln\left[\frac{N_{j\ell}}{N_\ell}\right]\\ &+\lambda_{i\ell}\left(D_\ell'(N_\ell) - \alpha_\ell c_i'(N_i) - \frac{1}{\theta_\ell}\left(\frac{N_\ell - N_{i\ell}}{N_\ell N_{i\ell}}\right)\right) - \sum_{k\neq\ell}\lambda_{ik}(\alpha_k c_i'(N_i)) + \sum_{j\in S_U}\lambda_{j\ell}\left(D_\ell'(N_\ell) + \frac{1}{\theta_\ell}\frac{1}{N_\ell}\right)\\ = &D_\ell(N_\ell) - \sum_{k=1}^{K}\alpha_k N_{ik}c_i'(N_i) - \alpha_\ell c_j(N_j) - \frac{1}{\theta_\ell}\ln\left[\frac{N_{j\ell}}{N_\ell}\right]\\ &+\lambda_{i\ell}\left(D_\ell'(N_\ell) - \frac{1}{\theta_\ell}\left(\frac{N_\ell - N_{i\ell}}{N_\ell N_{i\ell}}\right)\right) - \sum_{k}\lambda_{ik}\alpha_k c_i'(N_i)\\ &+\sum_{j\in S_U}\lambda_{j\ell}\left(D_\ell'(N_\ell) + \frac{1}{\theta_\ell}\frac{1}{N_\ell}\right) = 0, \forall i\in S_T, \ell = 1...K\end{aligned} \tag{C.2}$$

For the untolled route the first-order condition is given by:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial N_{i\ell}} = &D_\ell(N_\ell) - \sum_{k=1}^{K}\alpha_k N_{ik}c_i'(N_i) - \alpha_\ell c_j(N_j) - \frac{1}{\theta_\ell}\ln\left[\frac{N_{j\ell}}{N_\ell}\right]\\ &+\lambda_{i\ell}\left(D_\ell'(N_\ell) - \frac{1}{\theta_\ell}\left(\frac{N_\ell - N_{i\ell}}{N_\ell N_{i\ell}}\right)\right) - \sum_{k}\lambda_{ik}\alpha_k c_i'(N_i)\\ &+\sum_{j\in S_T}\lambda_{j\ell}\left(D_\ell'(N_\ell) + \frac{1}{\theta_\ell}\frac{1}{N_\ell}\right) = 0, \forall i\in S_U, \ell = 1...K.\end{aligned} \tag{C.3}$$

For the tolls and the multipliers we obtain:

$$\frac{\partial L}{\partial f_{i\ell}} = -\lambda_{i\ell} = 0, \forall i\in S_T, \ell = 1...K. \tag{C.4}$$

$$\frac{\partial L}{\partial \lambda_{il}} = D_\ell(N_\ell) - f_{i\ell} - \alpha_\ell c_i(N_i) - \frac{1}{\theta_\ell} \ln \left[\frac{N_{i\ell}}{N_\ell} \right] = 0, \forall i \in S_T, \forall \ell = 1...K. \tag{C.5}$$

$$\frac{\partial L}{\partial \lambda_{il}} = D_\ell(N_\ell) - \alpha_\ell c_i(N_i) - \frac{1}{\theta_\ell} \ln \left[\frac{N_{i\ell}}{N_\ell} \right] = 0, \forall i \in S_U, \forall \ell = 1...K. \tag{C.6}$$

Substituting C.4, C.5 and C.6 in C.2 and C.3 gives:

$$f_{i\ell} = \sum_{k=1}^{K} \alpha_k N_{ik} c_i'(N_i) - \sum_{j \in S_U} \lambda_{j\ell} \left(D_\ell'(N_\ell) + \frac{1}{\theta_\ell} \frac{1}{N_\ell} \right), \forall i \in S_T, \forall \ell = 1...K. \tag{C.7}$$

showing that we need the group-specific multipliers of the untolled routes to obtain a closed-form solution for the toll. Furthermore we have:

$$\sum_{k=1}^{K} \alpha_k N_{ik} c_i'(N_i) = \lambda_{i\ell} \left(D_\ell'(N_\ell) - \frac{1}{\theta_\ell} \left(\frac{N_\ell - N_{i\ell}}{N_\ell N_{i\ell}} \right) \right) - \sum_k \lambda_{ik} \alpha_k c_i'(N_i), \forall i \in S_U, \forall \ell = 1...K. \tag{C.8}$$

The solution for the group-specific Lagrangian multipliers for group ℓ on route i can be obtained using the system of K equations (C.8) for a given route i. These systems can be written in matrix notation:

$$A_i \lambda_i = b_i, \forall i \in S_U \tag{C.9}$$

where λ_i is the $K \times 1$ vector with unknown multipliers for route i, b is the $K \times 1$ vector with each element equal to $\sum_{k=1}^{K} \alpha_k N_{ik} c_i'(N_i)$, and A is the following $K \times K$ matrix:

$$A_i = \begin{pmatrix} D_1'(N_1) - \alpha_1 c_i'(N_i) - \frac{1}{\theta_1}\left(\frac{N_1 - N_{i1}}{N_1 N_{i1}}\right) & -\alpha_2 c_i'(N_i) & \cdots & -\alpha_K c_i'(N_i) \\ -\alpha_1 c_i'(N_i) & D_2'(N_2) - \alpha_2 c_i'(N_i) - \frac{1}{\theta_1}\left(\frac{N_2 - N_{i2}}{N_2 N_{i2}}\right) & \cdots & -\alpha_K c_i'(N_i) \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_1 c_i'(N_i) & -\alpha_1 c_i'(N_i) & \cdots & D_K'(N_K) - \alpha_K c_i'(N_i) - \frac{1}{\theta_1}\left(\frac{N_K - N_{iK}}{N_K N_{iK}}\right) \end{pmatrix}$$

The solution for the vector λ_i can be found by Cramers' rule. Let $A_{i\ell}(b_i)$ be the matrix A_i with column l replaced by the vector b_i. The solution for the lth Lagrangian multiplier is given by a ratio of determinants:

$$\lambda_{i\ell}^* = \frac{\det(A_{i\ell}(b_i))}{\det(A_i)}, \forall i \in S_U, \forall \ell = 1...K. \tag{C.10}$$

and therefore we need $\det(A_i) \neq 0$ to have a unique solution for the multipliers. Equation (C.10) can be made more explicit using analytical expressions for the determinants. Because the matrix A_i has many common elements, its determinant can be written in a tractable closed-form:

$$\det(A_i) = \prod_{k=1}^{K} \left[D_k'(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k} \right] - \sum_{k=1}^{K} \alpha_k c_i'(N_i) \prod_{\substack{m=1 \\ m \neq k}}^{K} \left[D_m'(N_m) - \frac{1}{\theta_m} \frac{N_i - N_{im}}{N_{im} N_m} \right]. \tag{C.11}$$

We can divide out the first product term in equation (C.11):

$$\det(A_i) = \underbrace{\prod_{k=1}^{K} \left[D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k} \right]}_{<0} \underbrace{\left(1 - \sum_{k=1}^{K} \frac{\alpha_k c'_i(N_i)}{D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k}} \right)}_{>0}. \tag{C.12}$$

Because $D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k} < 0, \forall k = 1...K$, the first part in equation (C.12) will be a product of negative numbers resulting in a number that is unequal to 0. Because $\alpha_k c'_i(N_i) > 0$, and $D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k} < 0$, the summation is over negative numbers resulting in a positive number for the part between large brackets. Therefore equation (C.12) is unequal to 0 and a unique solution for the Lagrangian multipliers exists. The solution (C.10) can be further investigated by using the following analytical expression for the determinant $\det(A_{i\ell}(b_i))$:

$$\det(A_{i\ell}(b_i)) = \sum_{k=1}^{K} \alpha_k N_{ik} c'_i(N_i) \prod_{\substack{r=1 \\ r \neq \ell}}^{K} \left[D'_r(N_r) - \frac{1}{\theta_r} \frac{N_i - N_{ir}}{N_{ir} N_l} \right]. \tag{C.13}$$

We have $\det(A_{i\ell}(b_i)) \neq 0$ implying that all the Lagrangian multipliers for route i have a unique non-zero value. Substituting equations (C.12) and (C.13) in equation (C.10) gives:

$$\lambda^*_{i\ell} = \frac{\sum_{k=1}^{K} \alpha_k N_{ik} c'_i(N_i) \prod_{\substack{r=1 \\ r \neq \ell}}^{K} \left[D'_r(N_r) - \frac{1}{\theta_r} \frac{N_i - N_{ir}}{N_{ir} N_r} \right]}{\prod_{k=1}^{K} \left[D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k} \right] \left(1 - \sum_{k=1}^{K} \frac{\alpha_k c'_i(N_i)}{D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k}} \right)}, \tag{C.14}$$

which can be rewritten as:

$$\lambda^*_{i\ell} = \sum_{k=1}^{K} \alpha_k N_{ik} c'_i(N_i) \frac{1}{\left(D'_l(N_l) - \frac{1}{\theta_l} \frac{N_i - N_{i\ell}}{N_{i\ell} N_l} \right) \left(1 - \sum_{k=1}^{K} \frac{\alpha_k c'_i(N_i)}{D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k}} \right)}. \tag{C.15}$$

Taking the ℓth term out of the summation this reduces to:

$$\lambda^*_{i\ell} = \sum_{k=1}^{K} \alpha_k N_{ik} c'_i(N_i) \frac{1}{D'_l(N_l) - \frac{1}{\theta_l} \frac{N_i - N_{i\ell}}{N_{i\ell} N_l} - \alpha_l c'_i(N_i) - c'_i(N_i) \sum_{\substack{k=1 \\ k \neq l}}^{K} \alpha_k \frac{D'_l(N_l) - \frac{1}{\theta_l} \frac{N_i - N_{i\ell}}{N_{i\ell} N_l}}{D'_k(N_k) - \frac{1}{\theta_k} \frac{N_i - N_{ik}}{N_{ik} N_k}}}. \tag{C.16}$$

This shows that the Lagrangian multipliers for each group are non-positive. Because tolls enter the constraints negatively, an increase in the group-specific toll $f_{i\ell}$ from 0 (which is the SB case under consideration) to a positive value will lead to higher welfare. Substituting equation (C.16) in equation (C.7) gives:

$$f_{i\ell} = \sum_{k=1}^{K} \alpha_k N_{ik} c_i'(N_i) - \sum_{j\in S_U} \left(\frac{\sum_{k=1}^{K} \alpha_k N_{jk} c_j'(N_j) \left(D_l'(N_l) + \frac{1}{\theta_l}\frac{1}{N_l} \right)}{D_l'(N_l) - \frac{1}{\theta_l}\frac{N_j - N_{jl}}{N_{jl} N_l} - \alpha_l c_l'(N_j) - c_j'(N_j) \sum_{\substack{k=1\\k\neq l}}^{K} \alpha_k \frac{D_l'(N_l) - \frac{1}{\theta_l}\frac{N_j - N_{jl}}{N_{jl}N_l}}{D_k'(N_k) - \frac{1}{\theta_k}\frac{N_j - N_{jk}}{N_{jk}N_k}}} \right),$$

$$\forall j \in S_T, \forall \ell = 1...K. \tag{C.17}$$

This completes the proof.

Appendix D. Calibration of the asymmetric route flows

If we want to calibrate the model in the no-toll case for given values of θ, we have observed number of travellers for both routes and the corresponding total number of travellers. We therefore also have the observed route probabilities which are functions of these. The inverse demand is assumed to be linear and is given by equation (36). In the no-toll equilibrium we have two conditions that need to be satisfied, since the marginal benefits should be equal to the generalised price. Assuming $\beta_T = \beta_U = \beta$ this results in:

$$\begin{aligned} \delta_1 - \delta_2(N_T + N_U) - \frac{1}{\theta}\ln\left[\frac{N_T}{N}\right] &= \kappa_T + \beta N_T. \\ \delta_1 - \delta_2(N_T + N_U) - \frac{1}{\theta}\ln\left[\frac{N_U}{N}\right] &= ASC_U + \kappa_U + \beta N_U, \end{aligned} \tag{D.1}$$

where ASC_U is the alternative specific constant for route U. Solving equation (D.1) for ASC_U gives:

$$ASC_U = \kappa_T - \kappa_U + \beta(N_T - N_U) + \frac{1}{\theta}\ln\left[\frac{N_T}{N_U}\right]. \tag{D.2}$$

We want to have N_T^0 and N_U^0 as the flows in deterministic user equilibrium, implying $\kappa_T + \beta N_T^0 = \kappa_U + \beta N_U^0 \implies \kappa_T - \kappa_U + \beta(N_T^0 - N_U^0) = 0$. Substituting in equation (D.2) gives:

$$ASC_U = \frac{1}{\theta}\ln\left[\frac{N_T^0}{N_U^0}\right]. \tag{D.3}$$

The symmetric case $N_T^0 = N_U^0$ is a special case and gives $ASC_U = 0$. This completes the calibration for asymmetric route flows.