Welfare Effects of Congestion Pricing and Transit Services in Multiclass Multimodal Networks

Yang Liu and Yu (Marco) Nie

This paper analyzes the welfare effects of congestion pricing with the use of a general bimodal network with heterogeneous users. Transit service is modeled as a cheaper (because of its lower operating cost) and undesirable (because of the relatively long travel time) alternative to the highway network and is added to an origin–destination pair as an exclusive link. The analysis characterizes the critical users—those who suffer the greatest loss from pricing—as those who experience the least change in travel time after pricing. Accordingly, the paper shows that critical users could be rich, poor, or middle class, depending on their origin and destination. This finding highlights the spatial heterogeneity of distributional effects. Furthermore, numerical experiments indicate that (a) those with a low value of time tend to benefit more from greater coverage of transit services than those with intermediate values of time; (b) in the presence of poor transit coverage, users with access to transit may share a disproportionally greater burden for the congestion relief generated by an optimum toll for the system; and (c) the optimum toll for the system generally leads to a welfare gap between the rich and the poor larger than those of toll schemes that have proportionally lower magnitudes and achieve smaller efficiency improvements.

It is well known that public ownership of roads results in efficiency loss in the form of excessive congestion. Such a loss can be avoided if each individual is informed of the true travel cost through either a market mechanism, for example, privatization of the roads, as suggested previously (1), or a mechanism that allows the government to toll the roads (2). Pigou’s design for road pricing (2) has become a subject of extensive research investigation (3–6), and today, it is practiced in various forms around the world.

Although many economists advocate road pricing as a general transportation policy, politicians and the public have always considered it with skepticism. First, road pricing appears to be just another tax on an otherwise free public good. Second, and more important, a widespread belief is that congestion pricing is a regressive policy (7–9) because an individual with a lower value of time (VOT) always gains less from the travel time savings associated with pricing. This observation of road pricing does not imply that individuals’ welfare always changes monotonically with VOT. For example, Verhoef and Small have shown, using a simple two-link network, that a user’s loss (or gain) in VOT monotonically decreases (or increases) if marginal tolls are collected on all links (first-best pricing) (10). In the case of second-best pricing (that is, the case in which only a subset of links can be tolled), the user who suffers the highest welfare loss has an intermediate VOT and is indifferent between tolled and nontolled routes. Liu et al. (11) and Nie and Liu (12) examined the distributional effects of pricing on users with continuously distributed VOTs, using a two-link bimodal network. They noted that the user who is indifferent between the two modes (highway and transit) after pricing suffers the greatest loss. A bottleneck model with two parallel routes (of which only one is tolled) reveals similar results, namely, that users receiving the least welfare gain are those with an intermediate value of schedule delay and the lowest VOT corresponding to that value of schedule delay (13).

An understanding of the nonmonotonic pattern of distributional effects is important to the design of strategies that may be used to overcome uneven distributional effects. For example, if the welfare loss of the users who suffer from pricing the most can be identified, they can be compensated with a refund equal to their loss. Such a simple strategy would ensure that nobody’s welfare decreases; that is, the policy is Pareto improving. Whether the toll revenue would be enough to support such a refunding program is a different question and will not be discussed in this paper. The reader is referred to previously published work for more details (11, 12, 14).

Most existing analyses of welfare effects use models with simple network structures. Although these models are capable of revealing the basic properties of the problems, they often ignore some important aspects in the representation of real-world problems. To fill this gap, an analysis of the welfare effects of congestion pricing is performed by use of a general multiclass bimodal network model. The analysis shows that the users who suffer the greatest loss could be those with the highest, the lowest, or an intermediate VOT, depending on their origin and their destination. This is essentially because origin–destination (O-D) travel time may not always decrease from no-toll to tolled equilibrium even if pricing does reduce the total travel time on the system. Such complicated patterns of distributional effects, which have not been manifested in simple models, highlight the spatial heterogeneity of those effects.

In the multiclass bimodal network model described here, transit service is modeled as a cheaper (because of its lower operating cost) and undesirable (because of a travel time longer than that for highway routes used) alternative to the highway network and is added to an O-D pair as an exclusive link. The paper first discusses the solution structure of the multiclass bimodal network equilibrium problem, on the basis of which the welfare effects of congestion pricing are analyzed. With the help of a research tool called Talex, which can

Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208. Corresponding author: Y. Nie, y-nie@northwestern.edu

Transportation Research Record: Journal of the Transportation Research Board, No. 2283, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 34–43. DOI: 10.3141/2283-04
be downloaded for free at http://translab.civil.northwestern.edu/ nutrend_install/TalexSetup.msi, a numerical study is also conducted to examine the impacts of spatial coverage of transit services and the magnitude of tolls. Dial’s Algorithm B (15), provided in Talex, which belongs to a class of highly efficient bus-based traffic assignment algorithms, is adopted; and a few modest changes are made so that it may be applied to the multiclass model with transit links.

The next section introduces the multiclass bimodal network model and makes necessary definitions; the analytical properties of the equilibrium solution to the model and welfare effects of congestion pricing are then discussed. The implementation details needed to conduct a computational study are provided, and numerical results to verify the analyses are presented. The paper concludes with a summary of findings.

MODEL SETTINGS

Consider a general network \(G(N, A)\), where \(N\) represents a set of nodes and \(A\) represents a set of links. Let \(x_a\) and \(t_a(x_a)\) be the total link flow and travel time on link \(a \in A\), respectively, where \(t_a(\cdot)\) is a positive, separable, and strictly increasing function. Users are classified into groups according to their value of time \(\beta_m\), which is labeled such that \(0 < \beta_1 < \beta_2 < \cdots < \beta_M\). Let \(R\) and \(S\) denote the sets of origin and destination nodes, respectively. The trip rate (demand) for two reasons. First, corner solutions, in which all users between O-D pair prefer one mode to the other, may arise. Second, in the current setting, it is likely that none of the existing discrete VOTs satisfies the indifferent condition exactly. To formalize these discussions, the following lemma is first needed:

Lemma 1. At NTE, all highway users between the same O-D pair prefer one mode to the other, may arise. Second, in the current setting, it is likely that none of the existing discrete VOTs satisfies the indifferent condition exactly. To formalize these discussions, the following lemma is first needed:

Solution Structure

For each O-D pair, an indifferent class whose VOT is such that any user in this class has exactly the same travel cost on both modes may exist. The VOTs of the indifferent class at NTE and TUE are denoted \(\beta_m\) and \(\beta_m^{*}\), respectively. An indifferent class does not always exist, for two reasons. First, corner solutions, in which all users between an O-D pair prefer one mode to the other, may arise. Second, in the current setting, it is likely that none of the existing discrete VOTs satisfies the indifferent condition exactly. To formalize these discussions, the following lemma is first needed:

Lemma 1. At NTE, all highway users between the same O-D pair prefer one mode to the other, may arise. Second, in the current setting, it is likely that none of the existing discrete VOTs satisfies the indifferent condition exactly. To formalize these discussions, the following lemma is first needed:

As illustrated in Figure 1a, at NTE, all highway paths between an O-D pair \(rs\) that are used have the same travel time \(t_{rs}\) (according
to Lemma 1) and operating cost \( C_{rs} \) (in Figure 1, each path used is represented by a filled circle, and each unused path is indicated by an unfilled circle), whereas the transit line’s travel time and operating cost are \( \gamma_{rs} \) and \( oT_{rs} \), respectively (represented by a shaded circle). The slope of the line that connects the filled circle and the shaded circle gives the VOT of the indifferent class, that is,

\[
\beta_{rs} = \frac{o_T - o_T}{\gamma_{rs} - \gamma_{rs}} = \frac{\Delta_o}{\gamma_{rs} - \gamma_{rs}}. 
\]

As shown in Figure 1a, the cost of the user with VOT \( \beta \) who uses highway (transit) is the \( y \)-intercept of the line emitted from the highway (transit) circle (with slope \( \beta \)). Figure 1 shows that when \( \beta < \beta_{rs} \), the travel cost, including the cost of travel time and the out-of-pocket cost, on transit is always lower and that the opposite is true for \( \beta > \beta_{rs} \). In other words, users with a VOT of \( \beta < \beta_{rs} \) would prefer transit and those with a VOT of \( \beta > \beta_{rs} \) would prefer the highway. It follows that the corner solution occurs when \( \beta_1 > \beta_{rs} \), in which case all users will use highway, and when \( \beta_M < \beta_{rs} \), in which case all users will use transit.

Figure 1b illustrates the structure of the solution at TUE for O-D pair \( rs \). Optimal paths used by different classes may intuitively have different combinations of travel time and cost. However, as shown previously (16, 17), only the paths that are on the Pareto frontier (i.e., the lower envelop) may be used, because any path above the frontier is dominated, in the sense that at least one path that has both a lower cost and a shorter travel time always exists (and is thus unattractive to any user, regardless of users' VOTs). In this problem setting, the transit line must be a nondominated alternative and located at the right bottom of the frontier, because the out-of-pocket cost on the transit line must be lower than that on any highway path. Hence, the VOT of the indifferent user at TUE is given by the slope of the

---

**FIGURE 1** Equilibrium solutions for multiclass assignment problem with transit (O-D pair rs): (a) NTE, (b) TUE, (c) \( \beta_{rs} < \beta_{rs} \), and (d) \( \beta_{rs} > \beta_{rs} \).
line that connects the transit (the shaded circle) to the lowest-cost non-dominated highway path (the rightmost filled circle on the frontier), denoted \(1_c\):

\[
\beta^*_n = \frac{v^*_n + \Delta_n}{y_n - c_n}
\]

(8)

where \(v^*_n\) is the total toll on path \(1_c\). Similar to NTE, the plot shows that any user with VOT \(\beta > \beta^*_n\) always uses highway, and any user with \(\beta < \beta^*_n\) always takes transit. In fact, the observation can be generalized to all highway paths used.

Lemma 2. Start with the rightmost path and let the paths on the Pareto frontier be numbered from 1 to \(p\). Let \(\alpha_i\) be the slope of the line connecting path \(i\) to path \(i + 1\) on the frontier (Figure 1b). (a) \(\alpha_1 < \alpha_2 < \ldots < \alpha_{c-1};\) (b) the user whose VOT \(\beta\) satisfies \(\alpha_{c-1} < \beta < \alpha_1\) would always use path \(i\) and she or he would be indifferent to paths \(i\) and \(i + 1\) and when \(\beta\) is equal to \(\alpha_{c-1};\) (c) those who use path \(k\) must have a VOT equal to or greater than the VOT of those who use path \(l\) for any \(k\) and \(l\) such that \(p \geq k > l > 1\).

The following result summarizes the previous discussions of the flow distribution among highway and transit.

Proposition 1. For an O-D pair \(rs\), class \(l\) uses highway if and only if \(\beta^*_l \geq \beta^*_n\) at NTE or \(\beta < \beta^*_l\) at TUE; class \(l\) uses transit if and only if \(\beta^*_l \leq \beta^*_n\) at NTE or \(\beta \geq \beta^*_l\) at TUE.

Welfare Effects

To analyze the welfare effects of pricing, the focus here is on the O-D pairs for which a highway is used at NTE. If users always use transit, they will not be affected at all by pricing. To simplify the analysis, the transit travel time function \(y_t\) is assumed to be constant in this section. Relaxation of this assumption would complicate the analysis but would not change the nature of the main conclusions.

The goal is to find which users suffer from pricing the most, and this class is called the “critical class” in this paper. The critical class is defined as follows:

Definition 1. The critical class at TUE, which is the class of users between O-D pair \(rs\) subject to the highest increase in travel cost, is called the “critical class,” and the VOT for this class is called “critical VOT \(\beta^c\).”

This class is of interest partly because nobody should be worse off. Such a situation will not occur, as long as all users in the critical class are properly compensated with a uniform refund of toll revenues between each O-D pair. The following result indicates that users with higher VOTs benefit more from pricing when the travel time that they experience decreases.

Lemma 3. Consider two users with different VOTs. The user with a higher VOT always has an increase in travel cost because of pricing equal to or less than (more than) that of the other user if the travel times of both users at TUE are shorter (longer) than those at NTE.

Proof. See Appendix A.

The welfare effects of a pricing scheme on two-mode networks with multiclass users may now be analyzed. Two possible cases exist: \(\beta^*_n < \beta^*_l\) (Case 1) and \(\beta^*_n \geq \beta^*_l\) (Case 2). The analysis first shows that toll users are either better off or break even (their travel costs do not change) in Case 1. Therefore, Case 1 need not be of concern here because the pricing scheme accomplishes a Pareto-improving outcome on its own. The focus in Case 2, in which transit ridership is promoted, is to show that any user could become a critical class.

First consider \(\beta^*_n < \beta^*_l\). As illustrated in Figure 1c, \(\beta^*_n < \beta^*_l\) implies that all highway paths used must have a travel time shorter than \(\bar{c}_e\). According to Proposition 1, all users with \(\beta\) such that \(\beta^*_n < \beta < \beta^*_l\) will switch from transit to highway. They are not worse off since they can at least break even by staying on transit. According to Lemma 3, the users with a VOT of \(\beta \geq \beta^*_l\) should have a cost increase equal to or less than that of users with \(\beta^*_n < \beta < \beta^*_l\), which means that their benefit is equal to or greater than that of the latter users. To summarize, when \(\beta^*_n < \beta^*_l\), all highway users at TUE are better off and those who remain on transit break even.

In the case of \(\beta^*_n = \beta^*_l\), some highway paths that are used could have travel times longer than \(\bar{c}_e\), as shown in Figure 1d. In this case, all users with \(\beta\) such that \(\beta^*_n < \beta < \beta^*_l\) will be tolled off the highway and suffer a welfare loss. The following result shows how the critical class can be determined on the basis of the path travel time at NTE and TUE.

Proposition 2. Given a toll scheme on a two-mode network, the alternatives include a set of highway paths \(\{1, 2, \ldots, p\}\) (see Figure 1d) and transit, if it is available. The critical VOT between the O-D pair is

- The lowest VOT, if transit is not available or not used and travel time increases on all the nondominated paths, \(\beta^*_n > \beta^*_1 > \beta^*_2 > \ldots > \beta^*_n\).
- The highest VOT, if travel time increases on all the nondominated paths, \(\beta^*_n > \beta^*_1 > \beta^*_2 > \ldots > \beta^*_n\).
- Either the highest VOT among all users of path \(k\) or the lowest VOT among all users of path \(k + 1\), if \(\gamma_k > \gamma_{k+1} > \gamma_{k+2} > \ldots > \gamma_n\) or \(\gamma_k > \gamma_{k+1} > \gamma_{k+2} > \ldots > \gamma_n\).

Proof. The results follow from Lemma 3 and Lemma 2. For example, in Case 1 of Proposition 2, when all highway paths used have a travel time shorter than that at NTE, according to Lemma 3, users with higher VOTs always benefit more than users with lower VOTs. Therefore, the critical VOT is the lowest VOT. Case 2 and Case 3 can be derived in a similar way.

Remark 1. The results presented above suggest that highway users whose path travel time experienced at TUE is closer to that experienced at NTE generally suffer a greater welfare loss and that users in the critical class are always among those whose pricing-induced change in travel time is the smallest. In Case 3, if a class that is indifferent to paths \(k\) and \(k + 1\) exists, that class is the critical class. This result is consistent with observations from the two-route model of Verhoef and Small (10). When all travel times on the highway path are reduced, \(\gamma_k > \gamma_{k+1} > \gamma_{k+2} > \ldots > \gamma_n\), the class with either the highest VOT on transit or the lowest VOT on highway suffers the greatest loss. Again, if the indifferent class exists, it is the critical class.

Remark 2. The class with the lowest VOT may be critical only if transit is not used or not available (Case 1 in Proposition 2). However, when transit is a competitive option for travelers, users with either an intermediate VOT or the highest VOT would be the critical class. Thus, provision of a competitive transit alternative is likely to favor the poor, according to welfare loss, at the expense of the middle class.
Remark 3. The transit travel time is assumed to be constant in this paper. Relaxation of this assumption will make transit less attractive, but as long as the externality on highway is higher than that on transit, which is often observed in the real world, some users would be tolled off and transit travel time will increase. Therefore, those who stay at transit regardless of toll, will also be worse off. However, consideration of a flow-dependent cost function of transit will not affect the conclusion in Proposition 2 and will be further examined in another paper.

IMPLEMENTATION ISSUES

This section provides the implementation details needed to conduct a computational study of the welfare effects of congestion pricing in a multiclass bimodal network.

Choice of Pricing Schemes

Marginal-cost pricing is used as the benchmark scheme in the computational study because it maximizes social welfare (and minimizes system cost). Specifically, the following system-optimum (SO) problem is first solved:

\[
\min \sum_{m} \sum_{a} \left( f_{m} \lambda_{a}(x_{a}) x_{a}^{m} \right) + \sum_{m} \sum_{a} \left[ \left( \int_{0}^{x_{a}^{m}} \gamma_{a}(w) dw + q_{a}^{m} \alpha_{a}^{c} \right) \right] 
\]

subject to Constraints 2 to 4

The optimal link toll is then set as \( u_{a}^{s} = t(x_{a}^{s}) \sum_{m} \lambda_{a} x_{a}^{m} \) where \( x_{a}^{m} \) and \( x_{a}^{m} \) are total and class-specific link flows at SO, respectively. Because Equation 9 is not generally convex for \( x_{a}^{m} \) (18–20), detection of a global optimum can be difficult. In other words, a decent algorithm could be stuck at a local optimum, and the result could be a system cost worse than the actual minimum. Nevertheless, as long as an SO solution generates a pricing scheme that offers a meaningful reduction in system cost (improvement in social welfare) compared with that of NTE, whether it is a global optimum is less of a concern for the present purpose.

Solution Algorithm

All multiclass traffic assignment models discussed above, tolled and no-toll user equilibrium as well as cost-based SO, are solved by use of Dial’s Algorithm B (15), which belongs to a class of highly efficient bush-based assignment algorithms. To apply the algorithm in the multiclass case with transit links, a few modest changes are needed.

First, in the multiclass implementation, each class associated with the same origin has its own bush, which is equilibrated according to its own VOT. Second, the original network must be expanded to accommodate transit links. As illustrated in Figure 2, if a transit link exists for an O-D pair, a dummy node and a dummy link will be added to both the origin and the destination and the transit link will be created to connect the dummy origin and destination. The transit link will have a travel time function, \( \gamma_{a}(q_{a}) \), and an operating cost, \( \alpha_{a}^{c} \), whereas the highway operating cost \( \alpha_{a}^{h} \) will be added as a cost to the dummy origin link. The travel time on both origin and destination dummy links is 0. To reduce the size of the expanded network, all O-D pairs originating from the same origin are assumed to share the same dummy origin link and thus have identical highway operating costs. Because the analysis excludes the destination choice, the impact of this simplification on the results is not a major concern. Moreover, the setting described above implies that the highway operating cost (not including toll) would not affect the route choice. Although this may not be the case in reality, the focus here is the impact of operating costs on mode choice.

Issue of Nonuniqueness

It is well known that class-specific flows are not unique in multiclass traffic assignment models. For each class, the path flows are also not unique. Furthermore, although the total system travel time and travel cost at both NTE and TUE are unique, the system time and the system cost associated with each O-D pair at TUE are not. The issue of nonuniqueness is more of a problem in analyses that require quantities aggregated over class or O-D pairs. For example, if the analyst wants to know how the percentage of transit users from a particular class is affected by pricing, she or he is likely to obtain different answers from different assignment tools or the same tool with different algorithmic parameters. The reader is referred to the work of Boyce et al. for a comprehensive discussion of nonuniqueness and its implications for the practice of urban travel forecasting (21).

Additional conditions are typically required to determine class-specific link flows and path flows. One method is to use the entropy model:

\[
\max \sum_{m} \sum_{a} \left[ \left( \sum_{k} f_{m}^{a} \ln \frac{f_{m}^{a}}{d_{m}^{a}} \right) + q_{a}^{m} \ln \frac{d_{a}^{m}}{d_{a}^{m}} \right] 
\]

subject to

\( f \in \Omega^{s} \)

where \( \Omega^{s} \) equal to \( \{f| f \text{ satisfies Conditions 2 to 6}\} \). Lu and Nie showed that the entropy model (Equation 11) ensures the uniqueness of path flow solutions and stability (22). Bar-Gera showed that the solution to Equation 11 always satisfies the following proportionality condition, which is often described as a proportionality hypothesis because the condition entails a behavioral interpretation (23).
Let on the transit link is assumed to be a constant in these experiments. The travel time function is used to model link travel time. Transit links are added into this network as described above. For simplicity, the travel time function is used to model link travel time. Transit links are added into this network as the base on which to construct the multiclass bimodal network assignment problem is solved with the Talex research tool from H. Bar-Gera’s Transportation Network Test Problems web page (http://www.bgu.ac.il/bargera/tntp/). The network has 24 nodes, 76 links, and 528 O-D pairs. The Bureau of Public Roads function (http://www.bgu.ac.il/bargera/tntp/). The network has 24 nodes, 76 links, and 528 O-D pairs. The Bureau of Public Roads function

\[ t_C = t_0 (1 + 0.15(x/C))^x, \]

where \( t_0 \) is the free-flow travel time, \( x \) is the volume on the link, and \( C \) is the link capacity, is used to model link travel time. Transit links are added into this network as described above. For simplicity, the travel time function on the transit link is assumed to be a constant in these experiments. Let \( \Gamma_r \) denote the set of all O-D pairs that originate from \( r \) and have a transit connection, let \( c^v_{rs} \) be the minimal free-flow travel time between O-D pair \( rs \), let \( l^0_{rs} \) be the shortest distance between O-D pair \( rs \), and let \( \sigma_r \) and \( \sigma_C \) be the unit operating costs for transit and highway, respectively. The constant travel time on the transit line is

\[ \gamma_r = \xi_r c^v_{rs} \]  

where \( \xi_r \) is a positive scalar to ensure that transit is relatively slow compared with highway. The operating costs for transit and highway are calculated on the basis of the average shortest distance from origin \( r \) as follows:

![FIGURE 3](image)  
**FIGURE 3** Class-specific demand in discrete log-normal distribution.

Assumption 1 (proportionality). The distribution of vehicle flows on any pair of alternative route segments (PASs) should be identical among behaviorally similar travelers, regardless of their origin and destination.

Therefore, unique route flows may be solved by iterative checking and enforcement of proportionality on PASs (24). On the basis of this concept, a postprocessing algorithm that generates PASs and enforces the proportionality condition on them was implemented on top of Algorithm B. The algorithm borrows the basic idea from traffic assignment by paired alternative segments (23, 24) but was implemented independently.

**Numerical Experiments**

The numerical experiments use the Sioux Falls, South Dakota, network as the base on which to construct the multiclass bimodal problems. The data of the basic Sioux Falls problem can be obtained at H. Bar-Gera’s Transportation Network Test Problems web page (http://www.bgu.ac.il/bargera/tntp/). The network has 24 nodes, 76 links, and 528 O-D pairs. The Bureau of Public Roads function

\[ t_C = t_0 (1 + 0.15(x/C))^x, \]

where \( t_0 \) is the free-flow travel time, \( x \) is the volume on the link, and \( C \) is the link capacity, is used to model link travel time. Transit links are added into this network as described above. For simplicity, the travel time function on the transit link is assumed to be a constant in these experiments. Let \( \Gamma_r \) denote the set of all O-D pairs that originate from \( r \) and have a transit connection, let \( c^v_{rs} \) be the minimal free-flow travel time between O-D pair \( rs \), let \( l^0_{rs} \) be the shortest distance between O-D pair \( rs \), and let \( \sigma_r \) and \( \sigma_C \) be the unit operating costs for transit and highway, respectively. The constant travel time on the transit line is

\[ \gamma_r = \xi_r c^v_{rs} \]  

where \( \xi_r \) is a positive scalar to ensure that transit is relatively slow compared with highway. The operating costs for transit and highway are calculated on the basis of the average shortest distance from origin \( r \) as follows:

\[ o^r_C = \frac{\sigma_r}{\Gamma_r} \sum_{rs} l^0_{rs} \]  

\[ o^r_C = o^r_C \frac{\sigma_r}{\sigma_C} \]  

where \( |\Gamma_r| \) denotes the number of O-D pairs in the set \( \Gamma_r \). In the following experiments, \( \sigma_r \) is $0.30 per mile per passenger, \( \sigma_C \) ranges from $0.31 to $0.60 per mile per passenger, and \( \xi_r \) is set equal to 2. Equation 13 suggests that all O-D pairs having a transit line and originating from the same origin would share the same operating costs for both highway and transit, a result which is consistent with the network expansion described in Figure 2. For the O-D pairs that do not have transit, their highway operating cost is set equal to 0, because it would have no impact on equilibrium solutions.

VOT is assumed to follow a log-normal distribution in the population, with an average of $21/h and a variance of $110/h. These parameters are adopted from a study of commuters on State Route 91 (25). Users are discretized into 10 classes according to the log-normal curve, as shown in Figure 3. Finally, the same VOT distribution is applied to all O-D pairs. The multiclass bimodal network assignment problem is solved with the Talex research tool described above.

**Disaggregated Welfare Effects**

The welfare of individuals from different O-D pairs is first examined to verify the analytical results presented above. This experiment uses the Sioux Falls network with 24 transit lines. Six O-D pairs representing typical patterns of welfare effects from the SO toll are demonstrated in Figure 4. The following observations can be made from the plot:

- For the O-D pairs without transit service (O-D Pairs 10–1 and 1–7) or the O-D pairs whose transit service is never used (O-D Pair 10–15), the users’ loss because of pricing is linearly increasing or decreasing with their VOTs. A close look reveals that
  - Only one highway path is used between these O-D pairs at TUE and
The equilibrium travel time between O-D Pairs 1–7 and 10–15 decreases after addition of a toll and the travel time between O-D Pair 10–1 increases after addition of a toll. According to Proposition 2, the critical group for O-D Pairs 1–7 and 10–15 should have the lowest VOT, whereas for O-D Pair 10–1, the critical group should have the highest VOT. This agrees with the results shown in Figure 4.

- For the O-D pairs with competitive transit service:
  - The highway travel time is decreased and the indifferent class (Class 5) for O-D Pair 10–16 suffers the highest cost increase, represented by the vertex of the curve ($\beta_{10,16} = 0.59$).
  - No indifferent class exists between O-D Pair 11–14 (the indifferent class is empty), and the class with the lowest VOT on highway is critical, presented by the vertex of the curve for O-D Pair 11–14 ($\beta_{11,14} = 0.33$).
  - Because no time saving exists on the two highway paths for O-D Pair 17–19, the direct loss of users increases with VOTs and the richest class suffers the greatest loss.

Thus, the example presented above verifies that the critical group can be the group with the highest VOT, the lowest VOT, or an intermediate VOT.

### Aggregated Welfare Effects

In the experiment evaluating aggregated welfare effects, Sioux Falls networks with different configurations of transit lines were tested to examine how the density of transit lines would affect the distributional welfare impacts of pricing. Sort, in descending order, the 528 O-D pairs according to the loss of the critical classes after SO toll, and then select the first $N_t$ O-D pairs to introduce transit lines. The impacts of different transit configurations, identified by $N_t$, are plotted in Figures 5 and 6. Although each class between different O-D pairs may be affected differently by toll, that is, the spatial inequity issue referred to in the literature, the aggregated average impacts are shown here to focus on the overall impacts on heterogeneous users.
Two scenarios with different transit services are first considered: (a) all 528 O-D pairs have transit services, and (b) transit service is available only for 100 of the 528 O-D pairs. The SO tolls are first calculated for each scenario and applied to find tolled equilibrium. In this experiment, the operating cost parameter for highway ($\sigma$) is first set equal to $0.50/\text{mi}$. Figure 5 examines the impacts of SO toll on each class’s transit share in both scenarios. For the network with dense transit service (Figure 5a), after the SO toll is applied, the aggregated transit shares (y-axis) increase for classes with low VOTs (Classes 1 to 4 on the x-axis) and decrease for classes with high VOTs (Classes 5 to 10). The SO toll tends to help the rich population better access the highway by tolling the poor off the highway. For the network with sparser transit service (Figure 5b), those who do have the service tend to use it more (roughly 40% of the richest people use transit before toll), and their transit shares increase significantly after the SO toll.

The transit shares reported in Figure 5b are only for those users who have access to transit. In fact, after the toll, the majority of the users with access to transit use transit. Therefore, the SO toll probably places a harsher penalty on users with access to transit because the network is more effective in reducing congestion when they are priced out. These users likely share an unfairly greater burden for congestion relief produced by the SO toll. The transit share is relatively low when the service coverage is extensive. A possible explanation is that the existence of extensive transit lines provides significant congestion relief, which, ironically, makes transit services less attractive, on average, with no toll.

The results for all possible scenarios are shown next. In Figure 6a, the average welfare loss (or the increase in travel cost; y-axis) is plotted for each class (x-axis). The different shadings in Figure 6a represent the number of transit lines ($N_t$). When no more than 150 transit lines are present ($N_t \leq 150$), all classes are worse off. In general, an increase in accessibility to transit services improves all users’ welfare and the travelers with high VOTs benefit from pricing more than the travelers with low VOTs. Moreover, those with the lowest VOTs benefit more than those with intermediate VOTs when the coverage of transit services improves. For the sparser transit service ($N_t < 300$), the class with the lowest VOT (Class 1) is the worst off and the SO toll is strictly regressive; that is, a user’s benefit (or loss) increases (or decreases) with her or his value of time. For the network with dense transit service ($N_t \geq 300$), Class 1 does not suffer the greatest loss. Rather, Class 3 becomes the class that suffers the greatest loss.

A sensitivity analysis that tested several values of $\sigma$, ranging from $0.31$ to $0.60/\text{mi}$ was conducted. However, all tests generated similar findings, except that the average critical class is sometimes Class 2 instead of Class 3.

**Impacts of Magnitude of Toll**

How the magnitude of the toll would affect travelers’ welfare is now examined. To evaluate the effect, the experiment uses the network configuration with 528 transit lines described above and the operating cost parameter $\sigma$, is set equal to $0.60/\text{mi}$. The SO link toll is simply scaled down by a positive scalar between 0 and 1, which results in toll schemes that charge proportionally lower tolls but that are still able to improve both system time and system cost compared with NTE.

In Figure 6b, the average increase of travel costs (y-axis), inclusive of travel time, operation cost, and toll, are plotted for each class (x-axis). The shading of each bar represents the scale used to generate the toll scheme. Therefore, for each class in Figure 6b, the corresponding toll scheme increases from left to right. Users in Classes 5 to 10 always benefit, and users in Classes 2 to 4 are always worse off, regardless of the magnitude of the toll scheme. The class with the lowest VOT (Class 1) breaks even only because individuals in that class always use transit, regardless of toll. The closer that the scale is to 1 (SO toll), the larger that the welfare gap that the toll creates between the rich and the poor is. The latter observation highlights the trade-off between equity and efficiency,
which must be carefully balanced in the design and implementation of congestion pricing.

**SUMMARY OF FINDINGS**

It is well known that congestion pricing tends to favor rich people at the expense of the poor. This paper shows that the distributional effects of congestion pricing are complicated in general networks and are affected by, among other factors, the spatial distribution of heterogeneous travel demands, the availability of transit services, and the pricing schemes. The analyses indicate that the users who suffer the greatest loss are generally those who experience the smallest change in travel time after pricing. The implication of this characterization is that any user, the poorest, richest, or middle class, can become the biggest loser in the game. Such complexities, particularly the spatial heterogeneity of the distributional effects, must be taken into account when implementation strategies that aim to address equity (e.g., various refunding and compensation programs) are designed.

The numerical experiments investigate the impacts of the spatial coverage of transit services and the magnitude of toll on the welfare effects of congestion pricing. The main findings are summarized as follows:

- More transit services generally lead to higher user welfare gain (or lower welfare loss) after toll, if the cost required to build additional transit lines is not considered. Those with a low VOT tend to benefit more from the greater coverage of transit services than those with an intermediate VOT. The poorest often suffer from congestion pricing the most when transit service is limited, whereas the lower middle class becomes the biggest loser in the presence of extensive transit coverage.
- When the transit coverage is poor, users with access to transit may be punished more by an SO toll. In other words, they share a disproportionally greater burden for the congestion relief.
- The SO toll generally leads to a greater welfare gap between the rich and the poor compared with the gap in toll schemes that have proportionally lower magnitudes and achieve smaller improvements in efficiency. This finding highlights the importance of consideration of the trade-off between efficiency and equity when a pricing scheme is designed.

**ACKNOWLEDGMENTS**

Kenneth Small offered valuable comments on an earlier draft of this work. The authors also thank three anonymous reviewers for their editorial help.

**APPENDIX A**

Lemma 1

Proof. The NTE problem has a unique link flow solution. At equilibrium, the user on path \( k \) cannot find a path \( l \) with a shorter travel time \( c_l < c_k \), regardless of its VOT; otherwise, she or he can reduce the travel cost by switching to path \( l \) (in the lower bound of \( c_l < c_k \)), but it is not at equilibrium. Thus, the minimum travel time between a given O-D pair \( c_{rs} \), must be the same across user classes and used paths \( \bar{c}_{rs} = \bar{c}_{rs} = \bar{c}_{rs} \leq \sum \bar{f}_{rs} > 0, \sum \bar{f}_{rs} > 0 \).

**Lemma 3**

Proof. Consider the case in which the travel time of both users increases from NTE to TUE. Let \( \alpha_1 \) and \( \alpha_2 \) be the VOTs of the two users, identified as Users 1 and 2, and \( \alpha_1 < \alpha_2 \). The two users may or may not use the same path. If they do use the same path \( k \in \{1, \ldots, p\} \) (where \( k \) is one of the paths on the frontier; Figure 1b), the proof is simple because

\[
\Delta w_1 = \mu_1 + \alpha_1 (c_{rs} - \bar{c}_{rs}) > \mu_1 + \alpha_1 (c_{rs} - \bar{c}_{rs}) = \Delta w_2
\]

If the two users use two different paths, without a loss of generality, assume that User 1 uses path \( k \) and User 2 uses path \( k + 1 \). According to Lemma 2, this means

\[
\alpha_1 \mu_{rs}^{k+1} - \mu_k \leq \alpha_2 \mu_{rs}^{k+1} - \mu_k
\]

Use of the inequality presented above provides

\[
\Delta w_1 = \mu_1 + \alpha_1 (c_{rs} - \bar{c}_{rs}) \geq \mu_1 + \frac{\mu_{rs}^{k+1} - \mu_k}{c_{rs} - c_{rs}} (c_{rs} - \bar{c}_{rs})
\]

\[
\Delta w_2 = \mu_1 + \alpha_2 (c_{rs} - \bar{c}_{rs}) \leq \mu_1 + \frac{\mu_{rs}^{k+1} - \mu_k}{c_{rs} - c_{rs}} (c_{rs} - \bar{c}_{rs})
\]

The right-hand sides of the previous two inequalities are actually identical, so \( \Delta w_1 \geq \Delta w_2 \). The case in which the travel times of both users decreases from NTE to TUE can be proven similarly.

**REFERENCES**


*Any errors and mistakes are the authors’.*

*The Transportation Network Modeling Committee peer-reviewed this paper.*