Simulated Relationships between Highway Capacity, Transit Ridership, and Service Frequency

Robert B. Noland
Centre for Transport Studies, Imperial College, London

Abstract

This article analyzes the relationships between highway capacity additions and transit patronage, both in the short and long run. A methodology using a model of schedule disutility is shown to provide a technique to account for transit service frequency. This technique, combined with a supply-side model of a highway corridor is used to evaluate the impact of transit headway changes and highway capacity, increases on total transit ridership, using a synthetic sample of commuters. Simulation results are used to evaluate the impact on travel times and utility of the two modes and the long-run degradation of transit service predicted by the Downs-Thomson paradox.

While the results do not show congestion as necessarily being worse than before capacity expansion, they do show that transit service frequency could be reduced significantly over time.

Introduction

The relative inconvenience of transit service compared to single-occupant vehicles (SOVs) is often cited as one of the primary reasons that transit rider-
ship shares have been diminishing in recent years (U.S. Department of Transportation 1997). Much of this is due to new patterns of development that have decentralized jobs and other activities away from the urban core. This decentralization has resulted in difficulties in supplying transit services that provide coverage for the multitude of potential trips within a large metropolitan area.

Traditional transit services also run on fixed schedules with discrete time intervals. This creates an additional source of inconvenience for users, especially if the fixed schedule deviates significantly from one’s own desired schedule of activities. While frequent, more convenient service is difficult to provide in decentralized areas, service frequency has also been reduced in many urban areas and for trips to the central business district (CBD). It is well known that these service reductions will result in lower transit patronage (Voith 1991; Lago et al. 1981; Kain and Liu 1995). Morlok (1976) provided some of the first analyses demonstrating a relationship between transit frequency and passenger volumes.

Transit’s level of inconvenience can be defined in two different ways: spatial inconvenience and temporal inconvenience. Spatial inconvenience of transit is a function of changing urban settlement patterns and is driven by the decentralization of urban areas. Temporal inconvenience refers to transit service that is relatively infrequent on existing routes, whether it serves suburban destinations or traditional routes to the CBD. Temporal inconvenience and its interaction with highway capacity is the focus of this article.

Changes in the attributes of SOV travel also affect transit ridership, especially in the long run. For example, increased road capacity has resulted in greater convenience and access for motor vehicles and has certainly contributed to reductions in transit patronage.

Highway capacity increases tend to result in unforeseen consequences. One of the paradoxes of transportation is the Downs-Thomson effect. This effect hypothesizes that highway capacity improvements may actually increase overall congestion and travel times (Arnott and Small 1994). One of the immediate effects of a highway capacity expansion, for a given congested corridor,
is a shift from transit to private vehicle use by some travelers. The Downs-Thomson effect hypothesizes that this reduction in transit ridership will produce a reaction where either transit fares are raised to cover costs or service is reduced. This can occur for both privately operated systems that reduce service due to decreased revenue or for government-provided services that seek to minimize deficits for political reasons. Both reactions by the transit service provider tend to further diminish transit patronage and shift more people into private vehicles. In the worst-case scenario, transit service is completely eliminated and congestion within the corridor is worse than before the capacity expansion. Arnott and Small (1994) numerically show results where congestion can be worse after a highway capacity expansion. The procedure outlined in this article illustrates how reductions in service (and hence the convenience of transit) can result in this general effect, though the model used here does not show overall congestion increasing.

This topic has important implications for both the provision of transit services and how financing is efficiently provided. For example, Mohring (1972) suggested that one of the benefits of subsidizing transit service is to capture the external benefits of increased service frequency. Alternatively, Walters (1982) suggests that smaller vehicle sizes might be a more optimal solution that would enable more frequent service under competitive conditions. Voith (1991) makes a compelling argument for how increases in transit fares and service reductions (due to the need to reduce subsidies) actually lead to the need for increased subsidies as fewer people use the transit system. This article considers these effects by explicitly linking transit usage with changes in highway capacity, focusing on the relative scheduling convenience of the two modes.

The next section briefly discusses some issues and current practices in modeling transit and techniques used for modeling choice of travel time. This is followed by a discussion of the methodology used in this article as well as simulations that analyze alternative convenience levels and the Downs-Thomson effect. The conclusion provides some thoughts on interactions between transit and highway policy.
Current Modeling Practices

Most regional transportation planning in the United States utilizes some form of the four-step modeling process. For determining transit ridership, the key step is the mode choice model (usually a discrete choice logit model). These generally contain parameters related to cost, travel time (in and out of vehicle), and user demographics. There is normally no explicit attempt to account for the disutility due to scheduling effects. Out-of-vehicle travel times can serve as a proxy for some scheduling effects, though they may be more related to the reliability of the schedule. Generally, the coefficients on in-vehicle travel time are smaller than those on out-of-vehicle travel time. This probably implies some additional disutility associated with waiting, which is related to frequency of service.

When service frequency is high, waiting time may serve as a good proxy for scheduling effects. However, as Tisato (1998) points out, when service is less frequent, users will not arrive at transit stops randomly but will engage in "planned behavior" using information on transit departure times to better schedule their arrivals.

Recent research has attempted to model transfer penalties (Central Transportation Planning 1997), which could be interpreted as another form of inconvenience associated with transit. Transfer penalty coefficients were found to be significant and having a transfer was equated with about 15 to 20 minutes of travel time.

When transit service is unavailable between two zones within a region because it is very inconvenient, it will obviously not be modeled. Introduction of a new service between two previously unserved zones would be modeled using existing parameters estimated for the region or for a similar pair of zones.

Another branch of the literature is focused on approaches for optimizing transit system service parameters. These tend to assume fixed-passenger demand. The model developed by Spasovic and Schonfeld (1993) does not consider the impact of service frequency on overall passenger demand but does show how fixed-passenger demand leads to an optimal headway value. Banks (1990) develops a simulation model and concludes that optimal headways are
minimally affected by assuming fixed demand. Kocur and Hendrickson (1982) consider variable demand in their optimization of transit costs and user benefits. These methods for optimizing transit service (by minimizing operator and user costs) do not allow explicit analysis of highway expansion policies on transit service, which is one of the objectives of this article.

Rigid transit schedules are related to the timing and scheduling flexibility associated with trips. Small (1982) estimated a model of scheduling choice that provides a foundation for building time-period choice models. The model includes parameters for the disutility associated with not arriving at the desired time. These parameters have been defined as schedule delay-early (SDE) and schedule delay-late (SDL). Bates (1996) provides an extensive review of other time-period choice modeling efforts, but concludes that Small's overall approach is the most attractive.

Few if any of these approaches have been applied in general practice and schedule disutility has not been applied to transit. Cambridge Systematics (1997) provides an assessment of the current practice of time-of-day choice modeling with a review of some innovative approaches taken by metropolitan planning organizations. With a few exceptions, none of the approaches reviewed are true attempts at multivariate modeling of travel time choice. They generally attempt to provide additional detail on fractions of peak and off-peak travel by facility and by mode for various trip types. A few innovative approaches do apply a peak-spreading algorithm. Most attempts are somewhat limited in their ability to analyze policy variables that affect scheduling utility and the choice of travel time in conjunction with choice of mode.

The procedure outlined in the next section applies Small's schedule disutility model to analyze shifts between transit and highway usage within a simple hypothetical travel corridor. The impacts of scheduling and highway capacity expansion policies and their relative impact on transit usage can then be evaluated.

**Schedule Disutility and Transit Convenience**

The methodology developed in this article builds on previous work on schedule disutility by Small (1982) and applies it to a system with a fixed headway. Small (1982) used data collected in the San Francisco Bay area to esti-
mate a model of scheduling costs. The basic hypothesis is that commuters have a preferred time that they wish to arrive at work. They also want to minimize the time they spend traveling to work. It should generally be preferable to arrive before one's preferred time than to arrive later. A rational commuter will attempt to trade off between schedule delay and travel time to maximize utility. When there is no congested travel period, the trade-off is trivial and schedule delay is equal to zero. Under congested conditions, the commuter will choose a travel schedule that maximizes the utility between lengthier travel times and schedule delay. When applied to transit with a fixed headway, the commuter in some cases must choose to arrive either earlier or later than the preferred arrival time, if the transit schedule does not match the timing of the preferred arrival time. Small (1982) postulated the following general model:

\[ U = aT + \beta SDE + \gamma SDL + \theta D \]  

(1)

where:

- \( T \) = travel time,
- \( SDE \) = schedule delay-early, and
- \( SDL \) = schedule delay-late.

These are defined as:

\[ SDE = \begin{cases} SD & \text{if } SD > 0, \\ 0 & \text{otherwise} \end{cases} \] 
\[ (2) \]

\[ SDL = \begin{cases} -SD & \text{if } SD < 0, \\ 0 & \text{otherwise} \end{cases} \] 
\[ (3) \]

\( SD \) is total schedule delay or the difference between the actual and preferred arrival time. \( D \) is a dummy variable equal to 1 when \( SDL > 0 \) and would represent an additional fixed penalty for arriving even one minute late. Both \( SDE \) and \( SDL \) increase linearly as one arrives either earlier or later than the preferred arrival time.
Small (1982) estimated coefficients for this model using a disaggregate logit model. The coefficients derived were of the expected sign and relative magnitude; that is, $\beta > \alpha > \gamma$. Arriving early is less onerous than time spent traveling, which is less onerous than arriving late. All values were statistically significant. Small (1982) also analyzed other formulations using various demographic variables, whether the vehicle is a carpool or not, and models with a variable representing flexibility in workplace arrival times. The coefficients for the simple model (1) are:

$$U = -0.106T - 0.065SDE - 0.254SDL - 0.58D$$

This model can easily be applied to the case of a fixed-transit schedule. A commuter electing to use transit would generally have to choose some amount of early or late arrival even under uncongested conditions. This would simply be a function of how well the transit schedule matches the preferred work arrival time.

**Relationships between Fixed Headways and Schedule Disutility**

Assume that the scheduled arrival time of a transit vehicle is $t_0$. The transit schedule has a fixed headway between vehicles of $H$. Therefore, if the first transit vehicle arrives at scheduled time $t(1)$, the scheduled arrivals of all vehicles can be defined as:

$$t_1 = t(1)$$
$$t_2 = t(1) + H$$
$$t_3 = t(1) + 2H$$
....
$$t_S = t(1) + (S-1)H$$

In practice, $H$ may vary with time of day or even over the peak period. It is assumed here to be a fixed headway. It is also assumed that there is no uncertainty in length of headway so the problem of bunching of buses running with low headways in congested areas is ignored.
Schedule delay (SD) can be written as $SD = t_A - t_p$, where $t_A$ is the actual arrival time and $t_p$ is the preferred arrival time. The actual arrival time, $(t_A)$ is determined by the choice of home departure time $(t_h)$. $SDE$ and $SDL$ are redefined as functions of the preferred arrival time and the scheduled arrival, $t_S$ which for transit is equal to $t_A$:

$$SDL = \begin{cases} t_S - t_p, & \text{if } t_S - t_p > 0 \\ 0, & \text{otherwise} \end{cases}$$  \hspace{1cm} (6)

$$SDE = \begin{cases} t_p - t_S, & \text{if } t_S - t_p < 0 \\ 0, & \text{otherwise} \end{cases}$$  \hspace{1cm} (7)

This allows a more general specification of Equation 1 to be defined where the utility $(U)$ is a function of mode $(M)$, home departure time $(t_h)$, and preferred arrival time:

$$U(M, t_h, t_p) = \alpha T(M, t_h) + \beta SDE(M, t_h, t_p) + \gamma SDL(M, t_h, t_p) + \theta D(M, t_h, t_p)$$  \hspace{1cm} (8)

The volume of traffic $(V)$ that the traveler expects to encounter determines the choice of home departure time. This also implies a set arrival time $(t_A)$, which is a function of $t_h$. For the transit mode, one can assume that travel time is independent of congestion levels if the vehicles travel on a separate guide-way (e.g., a rail system).

The following section specifies a procedure for simulating the choice of both mode and departure time. This allows for the endogenization of actual vehicle travel times and provides a technique for measuring mode shifts for relative levels of temporal inconvenience.

**Simulation with Endogenous Congestion**

To simulate the impacts of various policies it is necessary to endogenize the impact of congestion on individual travelers. Vickrey (1969) originally specified a bottleneck model of congestion that Arnott et al. (1990) later com-
bined with a schedule disutility model to determine the impact of congestion-tolling policies. Their approach is aggregate and does not provide detail on individual traveler reactions. Chu (1993) developed an approach that provides disaggregate detail by incorporating a discrete choice model of scheduling (more detailed than Small's) with a model of congestion technology as specified by the U.S. Bureau of Public Roads (1964) and used previously by Henderson (1981, 1985). Noland (1997), Small et al. (1995), and Noland (1999) extend the Chu model to account for reliability of travel time.

A nested logit formulation can be used to model the choice of mode and the choice of schedule (or departure time). This is superior to using a simple multinomial logit specification by eliminating the problem of independence of irrelevant alternatives (Ben-Akiva and Lerman 1985). For example, a multinomial specification would be sensitive to the number of choices of transit departure times available. All else equal, this would by itself result in fewer individuals using the transit mode since any elimination of a given choice results in proportional increases in the use of all other choices. A nested logit structure avoids this problem.

Nested logit models specify a logsum term that is the logarithm of the sum of the utility of a given nest. In this case, the nest represents the choice of departure times, hence the logsum ($LS$) is defined as:

$$LS_M = \ln \sum_{i=1}^{k} \exp[U_i(M)]$$  \hspace{1cm} (9)

where: $U_i$ is the utility function defined in Equation 8, for a given mode ($M$), and the summation is over the $k$ choices of departure time, $t_h$.

The logsum is then used in the upper nest of the logit model:

$$P(t_h, M) = \frac{e^{U_M + \delta_M LS_M}}{\sum_M e^{U_M + \delta_M LS_M}}$$  \hspace{1cm} (10)

This allows the generation of choice probabilities, $P(t_h, M)$, for each
departure time and choice of mode. The coefficient for the logsum ($\delta_M$) used in the simulations that follow was borrowed from the model developed by Chu (1993). The value of the logsum for vehicles is chosen to be 0.6842 and for transit is 0.2242. $U_M$ is the utility of the chosen mode, which is limited to a single transit specific constant of -5.422 (again, derived from Chu 1993). Travel times for each mode are already contained in the lower nest and thus are already accounted for.

One could also use a more detailed mode choice specification that more fully describes the choice of transit. This could include alternative travel time parameters for the two modes. For simplicity, it is assumed that any additional disutility associated with transit is contained in the mode-specific constant.

The probabilistic choice demand model is applied using a synthetic sample of 5,000 individuals, each with a randomly assigned preferred arrival time, $t_p$. This is actually the time that an individual exits the highway facility. It is assumed that each commuter then faces some additional time to actually reach his or her desired location. The synthetic sample is drawn randomly from a normal distribution with mean preferred arrival time equal to 8:00 A.M. and standard deviation equal to 60 minutes. Sample enumeration of the synthetic sample allows the probabilistic demand model to forecast the probability of choosing specified departure times (for both the vehicle and transit mode), relative to the preferred arrival time for each individual.

To clearly measure the difference between a mode with fixed headways and one with maximum temporal convenience, the vehicle departure time choices are segmented into 121 1-minute choices of arrival times. Of these, 80 segments are for the choice of arriving between 1 minute and 80 minutes early. One choice is for arriving exactly at the preferred arrival time and 40 are for between 1 minute and 40 minutes late. The number of transit choices is determined by the specified headway for a given simulation. For example, if the headway is 5 minutes, then there will be 25 choices of scheduled transit service within the 121-minute time frame specified. A 10-minute headway would provide 13 choices (i.e., the number of transit choices = $120/H + 1$). Small's scheduling cost function (4) was estimated using 5-minute intervals over an hour for
arrival times between 42.5 minutes early and 17.5 minutes late. Interpolation of choices for smaller time segments is not completely unrealistic. It is assumed that the choices apply over the 2-hour range specified (rather than Small's 1-hour range), however, simulations using a 61-minute interval produce essentially the same qualitative results with minor quantitative changes.

Sample enumeration of the choice probabilities for each travel time segment is calculated relative to the individual's randomly assigned preferred arrival time. This distribution of relative departure times is then allocated to specific 10-minute travel time slots. For example, if one individual has a preferred arrival time of 8:35 A.M., the probability that a schedule delay is -20 minutes is equivalent to the probability that the individual arrives in the time interval between 8:10 and 8:20 A.M. Traffic volumes are calculated for specified 10-minute time intervals.

Once traffic volumes have been calculated for specific time slots, one can determine the impact on travel times. To do this, the supply model cited by the U.S. Bureau of Public Roads (1964) is used:

\[
T = l \left[ T^0 + T^\varepsilon \left( \frac{V}{C} \right)^\varepsilon \right] \tag{11}
\]

where:

- \( T \) = travel time in minutes,
- \( V \) = number of vehicles leaving the highway per hour,
- \( C \) = capacity of the facility,
- \( \varepsilon \) = elasticity parameter,
- \( l \) = length of the facility (assumed to be equal to five miles), and
- \( T^0 \) and \( T^\varepsilon \) = constants.

The values used here are the parameters from U.S. Bureau of Public Roads (1964): \( \varepsilon = 4 \) and \( T^\varepsilon/T^0 = 0.15 \). \( T^0 = 1.0 \) minute/mile represents a free-flow speed of 60 miles per hour (mph). Traffic volume, \( V \), is calculated at the point where the flow leaves the highway and is based on the expected work arrival time. This simulation methodology is adapted from Chu (1993).

Vol. 3, No. 1, 2000
The simulations modeled assume that transit is traveling on a fixed guideway and so would not be subject to congestion within the highway corridor. One could also simulate the model by placing transit vehicles (buses) on the highway and making adjustments to congested travel times including the buses in the traffic flow. This is not done in the simulations that follow so that transit speed is controlled as an exogenous variable.

New vehicle travel times are then fed back into the probabilistic choice model to determine a new distribution of departure times and mode choices. This process is continued until convergence conditions achieve a stable pattern of travel volumes over time. (Specifically, convergence is achieved when the sum of the absolute value of traffic volume differences between two iterations is less than one.) Simulation outputs include the congestion profile, the average travel delay, scheduling and mode choices, and the total cost (or utility).

**Results of Simulations**

This section discusses several simulations that were run to determine potential impacts on transit ridership. These include the ridership, travel time, and average utility effects of changes in transit headway and speed, and changes in highway capacity. Long-term responses of highway capacity are then evaluated by assuming transit operators will reduce service frequencies, as hypothesized by the Downs-Thomson effect.

**Impact of Headways on Transit Ridership**

A series of simulations were run to analyze the impact on demand for transit for varying transit headways, transit speed, and different highway capacity levels. It was found, not surprisingly, that as transit headways are increased (i.e., service frequency decreased), transit ridership volumes decline. Results are displayed in Figure 1 for a variety of capacity levels, transit speeds, and different headways.

The results show that decreasing headways (i.e., increasing convenience) is an effective policy for increasing transit ridership. This is, of course, based on the parameters used in the schedule disutility function of Equation 4; other functional forms could give somewhat different results, although the general effects should be similar.
Highway construction or expansion projects are often packaged with transit expansion projects (ostensibly to address environmental and/or equity concerns). These results suggest that if transit headways are reduced within a corridor that has a highway capacity expansion, there could be some additional shifting to transit. For example, Figure 1 shows that reducing headways from 20 minutes to 10 minutes while increasing capacity from 150 to 300 vehicles (per 10-minute interval) results in an increase in transit share. As will be shown, the optimal headway may actually be higher, making it difficult to maintain a policy of increased service frequency.

Figure 1. Transit volumes for differing levels of highway capacity, transit headway, and transit speed
The same basic relationship between headway and transit ridership is maintained for each of the simulations with different capacity and transit speed inputs. The output suggests a fairly simple relationship between transit headway and transit volumes. For this reason, a simple linear regression was analyzed relating transit volumes to headway, capacity, and transit travel times. These results are shown in Table 1 for a logarithmic transformation of the data. Not surprisingly, all estimated coefficients are statistically significant. The logarithmic transform allows one to read the parameter estimates as elasticity measures. Transit headway shows a relatively high elasticity value indicating that a 1 percent increase (decrease) in transit headways can reduce (increase) transit ridership by about 0.77 percent. Lago et al. (1981) measured headway elasticities that ranged from about -0.22 to -0.76 depending on various conditions. They found larger elasticities during off-peak periods when service levels were generally low and lower elasticities at the peak, probably reflecting the inability of those traveling at peak periods to reschedule their trips. The simulation modeled here makes no assumptions about individuals being transit captive, which would, of course, result in lower aggregate elasticity values.

<table>
<thead>
<tr>
<th>Dependent Variable = \log (Transit Volume)</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.93</td>
<td>0.07</td>
</tr>
<tr>
<td>Log (Capacity)</td>
<td>-0.35</td>
<td>0.03</td>
</tr>
<tr>
<td>Log (Headway)</td>
<td>-0.77</td>
<td>0.01</td>
</tr>
<tr>
<td>Log (Transit travel time)</td>
<td>-0.24</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Another recent study by Kain and Liu (1995) estimated elasticities of revenue miles. Their estimate represents a measure of service quality similar, but different, than a headway measure. Their elasticities ranged from about 0.7 to as high as 1.0. While the comparison is not strictly comparable to headway results, it falls within the general range of the results above.
Voith (1991) measured short- and long-run elasticities using the number of peak and off-peak trains as a proxy for service quality. His elasticity values for peak-hour trains are 0.14 in the short run and 0.36 in the long run. For off-peak trains, the values are higher: 0.74 in the short run and 1.89 in the long run. Those traveling at peak may be more constrained in their choice of alternatives, hence they have lower elasticity values than those traveling during off-peak periods.

The analysis shows a clear relationship between transit usage and the frequency of transit service (i.e., headways). Long-run impacts and the Downs-Thomson paradox are analyzed and discussed below.

**Variations in Travel Times and Utility**

Average utility values, travel times, and modal shares for vehicle users and for transit users can be calculated using simulation results. This information provides some insight into how capacity changes affect these outcomes.

Average transit travel times were simulated at four levels (15, 30, 45, and 60 mph) while free-flow highway travel times were assumed to be 60 mph. Simulations with 60 mph transit travel speeds provide consistently faster peak travel times for transit vehicles than for highway vehicles. Table 2 shows that the immediate impact of a capacity expansion is to reduce both transit usage and both average and "peak" vehicle travel times in all cases. Potential longer term travel time impacts are discussed below.

Of more interest than travel times is the impact on average total utility. Separate components of utility, such as travel time utility, schedule delay utility, and lateness penalty utilities, are calculated using the parameters of Equation 4. These results are shown in Table 3. Average utility per traveler increases as capacity is increased. This is even true for the case where transit speeds exceed vehicle speeds. The only component of utility for vehicle users that significantly changes is the utility associated with travel time. The schedule delay utilities do not vary with capacity or speed of transit service. The average utility for transit users also does not vary.

As transit headways are increased, the average utility for all travelers is expected to decrease as shown in Table 4. One would expect that this is pri-
## Table 2
Travel Time (simulations for headway = 5 minutes)

<table>
<thead>
<tr>
<th>Capacity (volume/10-minute interval)</th>
<th>Transit Speed (mph)</th>
<th>Number of Vehicles</th>
<th>Number of Transit Users</th>
<th>Average Travel Time (minutes)</th>
<th>Average Vehicle Travel Time (minutes)</th>
<th>Average Transit Travel Time (minutes)</th>
<th>Average Peak Travel Time (minutes)</th>
<th>Average Peak Vehicle Travel Time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>15</td>
<td>3,716.17</td>
<td>1,283.76</td>
<td>10.133</td>
<td>6.724</td>
<td>20.00</td>
<td>10.655</td>
<td>7.415</td>
</tr>
<tr>
<td>225</td>
<td>15</td>
<td>3,861.69</td>
<td>1,138.24</td>
<td>8.774</td>
<td>5.466</td>
<td>20.00</td>
<td>8.815</td>
<td>5.653</td>
</tr>
<tr>
<td>300</td>
<td>15</td>
<td>3,893.97</td>
<td>1,105.96</td>
<td>8.442</td>
<td>5.159</td>
<td>20.00</td>
<td>8.358</td>
<td>5.223</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>3,481.93</td>
<td>1,518.01</td>
<td>7.467</td>
<td>6.363</td>
<td>10.00</td>
<td>7.844</td>
<td>6.912</td>
</tr>
<tr>
<td>225</td>
<td>30</td>
<td>3,610.34</td>
<td>1,389.59</td>
<td>6.649</td>
<td>5.359</td>
<td>10.00</td>
<td>6.712</td>
<td>5.503</td>
</tr>
<tr>
<td>300</td>
<td>30</td>
<td>3,638.70</td>
<td>1,361.23</td>
<td>6.450</td>
<td>5.122</td>
<td>10.00</td>
<td>6.434</td>
<td>5.170</td>
</tr>
<tr>
<td>225</td>
<td>45</td>
<td>3,519.39</td>
<td>1,480.55</td>
<td>5.722</td>
<td>5.325</td>
<td>6.67</td>
<td>5.803</td>
<td>5.456</td>
</tr>
<tr>
<td>300</td>
<td>45</td>
<td>3,546.06</td>
<td>1,453.87</td>
<td>5.562</td>
<td>5.110</td>
<td>6.67</td>
<td>5.577</td>
<td>5.154</td>
</tr>
<tr>
<td>150</td>
<td>60</td>
<td>3,354.91</td>
<td>1,645.03</td>
<td>5.800</td>
<td>6.193</td>
<td>5.00</td>
<td>6.128</td>
<td>6.672</td>
</tr>
<tr>
<td>225</td>
<td>60</td>
<td>3,472.63</td>
<td>1,527.31</td>
<td>5.214</td>
<td>5.309</td>
<td>5.00</td>
<td>5.305</td>
<td>5.433</td>
</tr>
<tr>
<td>300</td>
<td>60</td>
<td>3,498.41</td>
<td>1,501.53</td>
<td>5.073</td>
<td>5.104</td>
<td>5.00</td>
<td>5.104</td>
<td>5.146</td>
</tr>
</tbody>
</table>
Table 3
Average Utility of Travel (simulations for headway = 5 minutes)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>15</td>
<td>-1.524</td>
<td>0.068</td>
<td>-6.130</td>
<td>-1.1137</td>
<td>-1.7072</td>
<td>-0.2177</td>
<td>-0.7373</td>
<td>-0.0795</td>
<td>-0.1248</td>
</tr>
<tr>
<td>225</td>
<td>15</td>
<td>-1.275</td>
<td>0.157</td>
<td>-6.133</td>
<td>-1.1083</td>
<td>-1.7073</td>
<td>-0.2166</td>
<td>-0.7374</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>300</td>
<td>15</td>
<td>-1.217</td>
<td>0.179</td>
<td>-6.133</td>
<td>-1.1071</td>
<td>-1.7073</td>
<td>-0.2163</td>
<td>-0.7374</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>150</td>
<td>30</td>
<td>-1.717</td>
<td>0.102</td>
<td>-5.890</td>
<td>-1.1120</td>
<td>-1.7071</td>
<td>-0.2173</td>
<td>-0.7373</td>
<td>-0.0794</td>
<td>-0.1248</td>
</tr>
<tr>
<td>225</td>
<td>30</td>
<td>-1.513</td>
<td>0.173</td>
<td>-5.891</td>
<td>-1.1079</td>
<td>-1.7071</td>
<td>-0.2165</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>300</td>
<td>30</td>
<td>-1.466</td>
<td>0.190</td>
<td>-5.892</td>
<td>-1.1070</td>
<td>-1.7071</td>
<td>-0.2163</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>150</td>
<td>45</td>
<td>-1.784</td>
<td>0.113</td>
<td>-5.810</td>
<td>-1.1115</td>
<td>-1.7071</td>
<td>-0.2172</td>
<td>-0.7373</td>
<td>-0.0794</td>
<td>-0.1248</td>
</tr>
<tr>
<td>225</td>
<td>45</td>
<td>-1.595</td>
<td>0.178</td>
<td>-5.811</td>
<td>-1.1077</td>
<td>-1.7071</td>
<td>-0.2165</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>300</td>
<td>45</td>
<td>-1.553</td>
<td>0.193</td>
<td>-5.811</td>
<td>-1.1069</td>
<td>-1.7070</td>
<td>-0.2163</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>150</td>
<td>60</td>
<td>-1.818</td>
<td>0.119</td>
<td>-5.770</td>
<td>-1.1112</td>
<td>-1.7071</td>
<td>-0.2172</td>
<td>-0.7373</td>
<td>-0.0794</td>
<td>-0.1248</td>
</tr>
<tr>
<td>225</td>
<td>60</td>
<td>-1.637</td>
<td>0.181</td>
<td>-5.771</td>
<td>-1.1077</td>
<td>-1.7070</td>
<td>-0.2165</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
<tr>
<td>300</td>
<td>60</td>
<td>-1.596</td>
<td>0.195</td>
<td>-5.771</td>
<td>-1.1069</td>
<td>-1.7070</td>
<td>-0.2163</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
</tr>
</tbody>
</table>
### Table 4
Average Utility of Travel for Different Transit Headways and Highway Capacity (speed = 30 mph)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>150</td>
<td>-1.717</td>
<td>0.102</td>
<td>-5.890</td>
<td>-1.120</td>
<td>-1.7071</td>
<td>-0.2173</td>
<td>-0.7373</td>
<td>-0.0794</td>
<td>-0.1248</td>
<td>-2.7231</td>
</tr>
<tr>
<td>10</td>
<td>150</td>
<td>-1.190</td>
<td>0.018</td>
<td>-6.052</td>
<td>-1.1164</td>
<td>-1.7947</td>
<td>-0.2182</td>
<td>-0.7209</td>
<td>-0.0795</td>
<td>-0.1222</td>
<td>-2.8426</td>
</tr>
<tr>
<td>20</td>
<td>150</td>
<td>-0.830</td>
<td>-0.059</td>
<td>-6.217</td>
<td>-1.1213</td>
<td>-1.9598</td>
<td>-0.2192</td>
<td>-0.6918</td>
<td>-0.0797</td>
<td>-0.1180</td>
<td>-2.9520</td>
</tr>
<tr>
<td>40</td>
<td>150</td>
<td>-0.598</td>
<td>-0.122</td>
<td>-6.391</td>
<td>-1.1259</td>
<td>-2.2465</td>
<td>-0.2201</td>
<td>-0.6484</td>
<td>-0.0798</td>
<td>-0.1099</td>
<td>-3.0407</td>
</tr>
<tr>
<td>5</td>
<td>300</td>
<td>-1.466</td>
<td>0.190</td>
<td>-5.892</td>
<td>-1.1070</td>
<td>-1.7071</td>
<td>-0.2163</td>
<td>-0.7373</td>
<td>-0.0793</td>
<td>-0.1248</td>
<td>-2.5985</td>
</tr>
<tr>
<td>10</td>
<td>300</td>
<td>-0.834</td>
<td>0.165</td>
<td>-6.054</td>
<td>-1.1074</td>
<td>-1.7945</td>
<td>-0.2164</td>
<td>-0.7209</td>
<td>-0.0793</td>
<td>-0.1222</td>
<td>-2.6344</td>
</tr>
<tr>
<td>20</td>
<td>300</td>
<td>-0.423</td>
<td>0.148</td>
<td>-6.219</td>
<td>-1.1077</td>
<td>-1.9597</td>
<td>-0.2165</td>
<td>-0.6920</td>
<td>-0.0793</td>
<td>-0.1179</td>
<td>-2.6597</td>
</tr>
<tr>
<td>40</td>
<td>300</td>
<td>-0.181</td>
<td>0.137</td>
<td>-6.395</td>
<td>-1.1080</td>
<td>-2.2457</td>
<td>-0.2165</td>
<td>-0.6488</td>
<td>-0.0793</td>
<td>-0.1095</td>
<td>-2.6756</td>
</tr>
</tbody>
</table>
marily due to the shift to vehicle travel resulting in some increase in congestion and/or schedule delay. Most of this decrease in utility falls on vehicle users and is driven mainly by increases in travel time (or decreases in travel time utility for vehicles). As transit headways increase to 40 minutes, average utility for transit users also decreases. This effect is driven by reductions in the components of utility associated with schedule delay with minor variation due to the time component.\(^8\)

When capacity is increased, average utility for all travelers does improve. The change is primarily due to a shift from transit use. Average utility for transit users stays constant as capacity increases (although the total number of transit users is less). Interestingly, for both vehicle and transit users, the components of scheduling utility do not vary significantly. This shift is driven by vehicle travel time reductions associated with capacity increases, which, as will be seen, are overestimated when long-term responses are not considered.

**Long-Term Responses: The Downs-Thomson Paradox**

Increases in highway capacity have long been known to attract additional traffic (Downs 1962). The immediate impact occurs due to rescheduling and route shifting but other impacts include the generation of previously avoided trips and shifts from transit to motor vehicles. The simulations clearly demonstrate this latter effect in combination with rescheduling of trips toward the peak.

One of the more perverse effects of adding highway capacity is the Downs-Thomson paradox (Arnott and Small 1994). This paradox describes how a highway capacity increase could actually increase total congestion. If the capacity increase occurs in a corridor served by transit, it could result in a reduction in transit service frequency shifting additional people to motor vehicles. In some cases this could increase total travel time within the corridor or at least diminish the originally planned benefits of expanding the facility.

The simulation results are used to estimate how a capacity expansion can lead to long-term degradation in transit service. Assume first that there is an initial increase in highway capacity. This results in a short-run decrease in transit ridership (as discussed previously and demonstrated by the simulations).

---

*Vol. 3, No. 1, 2000*
The Downs-Thomson paradox can then come into play. The reduction in transit ridership triggers either an increase in transit fares (to cover lost revenue) or a decrease in service frequency (to reduce costs). If transit ridership is reduced, for example by 10 percent, it is assumed that service frequency is reduced by 10 percent (headway increased by 10%). This leads to a further reduction in transit usage. The regression displayed in Table 1 is used to calculate this iterative effect until convergence is achieved. Results are shown in Table 5.

### Table 5
Changes in Headway due to Highway Capacity Increases

<table>
<thead>
<tr>
<th>New Capacity (volume/10-minute interval)</th>
<th>Optimal Headway (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>5.48</td>
</tr>
<tr>
<td>175</td>
<td>6.15</td>
</tr>
<tr>
<td>200</td>
<td>7.11</td>
</tr>
<tr>
<td>225</td>
<td>7.93</td>
</tr>
<tr>
<td>300</td>
<td>9.82</td>
</tr>
<tr>
<td>450</td>
<td>12.24</td>
</tr>
<tr>
<td>600</td>
<td>13.78</td>
</tr>
</tbody>
</table>

Note: Original capacity is 150 vehicles/10-minute interval. Original headway is 5 minutes.

Table 5 describes results assuming that headways are initially equal to 5 minutes and capacity is equal to 150 vehicles per 10-minute interval. If capacity is increased to 225 vehicles per 10-minute interval, then a new equilibrium will be established such that the optimal headway is now 7.93 minutes. An increase in capacity to 600 vehicles per 10-minute interval results in a new equilibrium at an optimal headway of 13.78 minutes. These results are not dependent on transit speed, though different transit speeds result in different volumes of transit ridership. Figure 2 graphs the optimal transit headway versus the increase in highway capacity. Initially, relative increases in optimal headway are rather large, diminishing as larger increases in capacity occur. This suggests that small increases in highway capacity can potentially result in
pressures for relatively large reductions in transit service frequency to obtain the optimal level of service. Figure 3 shows the difference in transit ridership between an initial equilibrium and a full equilibrium effect when transit headways are adjusted to a new optimal level. The effect is quite substantial and very large as capacity levels increase.

This clearly shows that long-term reductions in transit ridership can be induced by increases in highway capacity without any change in transit fares. The Downs-Thomson paradox implies that overall congestion levels could be worse than before the capacity expansion. In the examples analyzed here, this does not seem to be the case. The capacity increase still results in reductions in travel time even after the reduction in transit frequency. For example, the optimal headway after expanding highway capacity to 300 vehicles per 10-minute interval is 9.82 minutes (Table 5). Simulated average vehicle times for a capacity of 300 and a headway of 10 minutes are 5.25 minutes, still less than the average vehicle travel time of over 6 minutes (Table 2). Utility values are also still greater even after a new optimal headway is established.

Mohring (1972) developed a model to determine optimal urban bus subsidies. As part of that model, Mohring asserts and estimates a relationship between optimal service frequencies and demand for transit use. This is for-
mulated as a “square root rule,” where the optimal frequency is equivalent to the square root of bus usage. The results here show the same general relationship. Figure 4 graphs optimal hourly service frequency (60 minutes/optimal headway) versus the square root of optimal transit ridership (as estimated after correcting for the Downs-Thomson effect). In general, the relationship is linear indicating a correspondence between these calculations and the results derived by Mohring (1972).

One caveat to the simulations is that the sample of 5,000 individuals used is static. One would expect capacity increases to induce generation of some new trips (other than just shifts from transit). Also, over time one would expect exogenous growth in travel due to population growth. If transit frequencies do not increase in proportion (due perhaps to a political decision to provide less support to transit since it is carrying fewer people), then again overall travel times could be reduced compared to not adding additional highway capacity.
Conclusions

The analysis presented here shows that transit service reductions clearly result in reduced transit ridership. The simulations do this using only scheduling costs as defined in Equation 4. The methodology also demonstrates that highway capacity increases result in both an immediate reduction in transit use and potentially a long-run reduction based on the behavioral assumptions of the Downs-Thomson paradox. While the simulations analyzed here do not show highway congestion to be worse than before the capacity expansion, other input assumptions could result in this occurring.

The results presented here should not be interpreted as definitive. The models used were relatively simple and many other factors could be attributed to modal shifts. However, the schedule disutility formulation used is relatively robust, and while the magnitude of the relative impacts may not be exact, the overall directions of the various changes due to headway increases or capacity changes are intuitively correct.

These types of impacts certainly question whether increasing road capacity is a solution for congested corridors or regions. Increasing service frequen-
cy of transit (and/or reducing fares) could, in some cases, reduce vehicle travel. Despite the innovations of the U.S. Intermodal Surface Transportation Efficiency Act of 1991 and its successor, the Transportation Equity Act of 1998, federal funding does not contribute major funding to transit operations. Most funding is restricted to capital improvements. Better uses of “transit” money may be to increase service frequency (and/or reduce fares). Decision-makers at the state and federal levels should evaluate the ability of increased transit service (on existing routes) as a means of meeting both transportation and environmental goals.

Acknowledgements
The author wishes to thank Ken Small, Xuehao Chu, and two anonymous reviewers for providing suggestions on improving this article. The author takes full responsibility for any remaining errors or omissions. Portions of this article were developed while the author was employed at the U.S. Environmental Protection Agency (EPA). The views expressed do not necessarily represent the views of the U.S. EPA or the U.S. government. A previous version of this article was presented at the 78th Annual Meeting of the Transportation Research Board (paper number 991527).

Endnotes
1. Other transit services that are not considered traditional include jitneys and demand-activated services (Klein et al. 1997; Cervero 1996).
2. Or as is discussed in the simulations, the commuter will choose a mode other than transit that better matches their preferred schedule.
3. This model assumes no stochasticity in travel times from day to day. Noland and Small (1995) developed a model of uncertain travel times.
4. Capacity levels shown are based on 10-minute travel time intervals and were selected to provide realistic levels of congestion for a simulation using only 5,000 travelers. This was done primarily to shorten computational time.
5. The simulations are only assuming travel within a specified five-mile corridor. In any specific situation, one would expect additional door-to-door travel times to be associated with each mode.
6. The “peak” here is defined as trips arriving between 7:00 A.M. and 9:00 A.M.
7. Some minor variation in the average utilities is due, most likely, to rounding errors in the simulation. The values are certainly not significant to three decimal places.
8. While travel time for transit is modeled as constant, the average utility varies slightly due to changes in the logsum associated with alternative headways.

9. The iteration could also be calculated using the overall simulation approach, but this is computationally difficult due to the integer headway values used in the simulations.

References


**About the Author**

**Robert B. Noland** (r.noland@ic.ac.uk) is the Rees Jeffreys Lecturer in Transport and the Environment in the Department of Civil and Environmental Engineering at Imperial College of Science, Technology, and Medicine in London. Previously he was a policy analyst with the Office of Policy of the U.S. Environmental Protection Agency. His research has focused on the behavioral and environmental impacts of transportation infrastructure.