# Accounting for Multimodal System Performance in Benefit-Cost Analysis of Transit Investment

Daniel Brod, *Hickling Lewis Brod, Inc.*

Benefit-cost analysis, in the conventional planning and modeling paradigm, estimates benefits from transit rail investment as the consumer surplus (willingness-to-pay) from forecast trips. New studies indicate that this paradigm, as currently implemented, fails to capture a wide array of benefits, namely improved multimodal system performance in congested corridors, transit-oriented development benefits, and cross-sectoral resource savings. The economic theory predicting improved multimodal performance in congested corridors when the transit mode is improved is developed, the empirical evidence supporting that theory is described, and a method for refining the practice of benefit-cost analysis to account for the benefit of improved multimodal performance is proposed. In urban corridors served by highways and a high-capacity transit mode, peak travel times and the modal split of trips will, in general, be influenced by highway capacity, relative prices, and individual preferences. However, in congested urban corridors door-to-door journey times are observed to be nearly equal across modes, converging toward the journey time by the high-capacity transit mode. The convergence of travel times is predicted from microeconomic theory. Empirical evidence from a recent study of 14 urban corridors in the United States supports this theoretical finding. It is further found that reducing transit headways contributes to the modal convergence of travel times. The principal policy implication of these findings is that improving the peak-hour performance of the high-capacity transit mode will also yield peak-hour performance improvements on the highway mode. The convergence of travel times across modes would not, in general, be the outcome predicted by the conventional models that forecast modal splits and transit ridership, which, in turn, form the basis for the analysis of benefits from transit investment. The multimodal effect of transit investment, as evidenced by the convergence of journey times, should be explicitly accounted for in the analysis of benefits. This can be accomplished through the calibration of estimated modal constants so that the assignment of trips to the urban transportation network yields nearly equal door-to-door journey times in the relevant market segments.

The current practice of benefit-cost analysis as applied to transit investments follows the conventional planning paradigm. Total demand is forecast as trips between zones; forecast trips are allocated to modes by means of a modal choice model; and, typically, the benefits from the proposed transit investment are estimated as the willingness-to-pay for the trips taken plus the benefits of reduced congestion on the highways. Recent studies conducted for the Federal Transit Administration's Office of Policy (publication forthcoming) have identified three areas in which this model fails to capture the full array of benefits from transit investment.

First, there remains the issue of the interaction between transportation investment and land use. The planning paradigm described was used to justify numerous road projects by assuming, for instance, that an outlying area would be developed. Under this assumption build and no-build scenarios were compared and road projects were shown to display strong benefits. Of course, it was often doubtful that development in the outlying area would have occurred in the absence of the road project. Furthermore, the conventional paradigm does not adequately address the issues of whether the planned road actually contributed to net new development or whether the development was preferable to other development alternatives. In contrast to highways, the benefit-cost analysis of transit rail investments does not account for the transit-oriented development that would legitimately be associated with a "build" scenario. A refinement of methods is under way that incorporates interactive land use and transit development scenarios, hedonic pricing methods for valuing development alternatives, and stated preference methods that seek to indirectly gauge the benefits of transit-oriented development.

The second area of benefits not captured by benefit-cost analysis is cross-sectoral resource savings. The absence of transit will restrict the mobility of some users and may require an increase in resource use for medical and social services. Studies demonstrating these benefits have been conducted in the United Kingdom, and methods for incorporating them into benefit-cost analysis are being developed.

Finally, conventional benefit-cost analysis does not account for the multimodal interrelationships that are observed in congested urban corridors. Mogridge (*1*) has shown that in congested urban corridors, door-to-door journey times are nearly equal and tend to converge to the journey time of the high-capacity transit mode. New evidence confirming this finding has been documented in recent and ongoing studies in the United States (see Table 1).

## Triple Convergence or Travel Time Convergence?

Downs (*3*) discusses as a principle of traffic analysis the notion of "triple convergence," whereby peak-hour traffic speeds converge spatially (across the road network) in time and across modes. Under the triple convergence principle, an improvement in peak-hour travel conditions on high-capacity roadways "will immediately elicit a triple convergence response, which will soon restore congestion during peak periods, although those periods may now be shorter." The prospects for improving transportation performance through transit investment are no less gloomy. Downs states that a new fixed-rail public transit system should initially reduce peak-period traffic congestion, but "as soon as drivers realize that expressways now permit faster travel, many will converge . . . onto those expressways during peak periods."

However, in congested urban corridors the observed convergence of peak-hour, door-to-door journey times—by the highway and high-capacity transit modes—suggests that a different dynamic is at work. If the travel time convergence dynamic were in effect, it is anticipated that a carefully chosen fixed-rail investment would indeed yield an improvement in journey times by highway. In general, the convergence of journey times to the journey time by the transit mode implies that a change in the performance of transit will result in a change in the performance of highways.

The phenomenon of travel time convergence to the transit journey time has profound policy implications for the planning and allocation of funds for transportation in metropolitan areas. Furthermore, it enables the application of benefit-cost analysis methods to alternatives across different modes (i.e., highway and transit projects are more readily comparable insofar as the cross-modal

TABLE 1 Door-to-Door Travel Times for Peak Journeys (*2*)

| Corridor | Auto Mode (Minutes) | High-Capacity Mode (Minutes) |
|---|---|---|
| New York Queens-Manhattan | 63.9 | 64.4 |
| San Francisco Bay Bridge | 72.3 | 73.1 |
| Philadelphia Schuylkill Expressway | 48.4 | 52.5 |
| Chicago - Midway | 54.2 | 60.6 |
| Chicago - O'Hare | 53.9 | 59.3 |
| Pittsburgh Parkway East | 38.1 | 42.5 |
| Princeton - New York | 113.4 | 104.9 |
| Washington - I-270 | 71.9 | 67.4 |

impacts can be compared where the conditions for trip time convergence are found to exist).

## Modal Explorers

What explains the phenomenon of travel time convergence? One claim is that a dynamic relationship exists that parallels that of a multilane highway: speeds across lanes tend to be equal because some drivers are "explorers" who seek out the faster-moving lane, thus driving the system to an equilibrium speed shared by all lanes. By the same token, in congested urban corridors some travelers and commuters are explorers. They are not committed through circumstance or strong preference to either mode and they behave as occasional mode switchers. If the transit mode has a high-speed, line-haul segment, the door-to-door journey time by this mode will be relatively stable, and small shifts in ridership will not significantly affect the journey time by the transit mode. On the other hand, under congested conditions even a 0.5 percent increase in highway traffic volume in the peak period can have a major impact on journey times. Because the journey time by transit is stable and determined by the speed of the high-capacity mode, transit "paces" the performance of the urban transportation system in the congested corridor. The modal explorers, like exploring drivers on the multilane highway, serve to bring about an equilibrium speed across modes as they seek travel time advantages across modes.

## Travel Time Equilibrium and Modal Choice

Whereas travel time represents a dominant component of the cost of trips, the generally accepted models of modal choice and the assignment of trips to networks would not predict travel times to be equal. Rather, the theory behind current practice anticipates modal choice by individuals to be driven by income, car ownership, money price differentials, and modal preferences that account for nonmoney factors like convenience, seamless travel, and so forth. The persistence of equal, or nearly equal, travel times across modes in congested corridors suggests that current theory fails to correctly capture modal interrelationships in a multimodal system.

The following model presents the economic theory for consumer behavior under congestion and develops the conditions under which door-to-door trip time by highway converges to the trip time by the high-capacity transit mode. It further demonstrates how congestion promotes the modal explorer behavior. Empirical evidence supporting the convergence of trip times to the high-capacity mode in congested corridors is presented. In the concluding section of this paper a proposed modification to the practice of the benefit-cost analysis of transit rail investment is discussed to account for this multimodal effect.

## Theoretical Structure

The theory presented here follows the standard model from public economics of utility maximization under a budget constraint with an external effect. Consider an individual who derives utility from consuming $z$ units per week of a basket of commodities. To generate the income required to purchase the consumption good, the individual must take $x$ trips per week (say, five inbound and five outbound) from a residential area to a central business district. The individual derives disutility, however, from the amount of time spent traveling. Whereas disutility may be derived differently from different types of travel time (i.e., driving, riding, walking, waiting in congestion, etc.), for simplicity the individual is assumed to be indifferent between travel times of different types. The individual can choose to travel by one of two modes, highway or high-capacity transit, each of which has a money price associated with the trip.

If there are $I$ individuals, the utility maximization problem of the $i$th individual is expressed as follows:

$$\max u^i(z, t) \text{ such that } x_1^i p_1 + x_2^i p_2 + z \leq y^i \quad (1)$$

where $t$ represents time spent commuting and $x_1^i$ and $x_2^i$ are the number of trips taken by the highway and the transit modes, respectively. The prices $P_1$ and $P_2$ are the money cost of a trip by each mode. $y^i$ is the individual's income. The price of the consumption good $z$ is 1.

The utility function is assumed to be continuous and twice differentiable, having the following properties:

$$u_z^i > 0 \qquad u_{zz}^i < 0 \qquad u_t^i < 0 \qquad u_{tt}^i < 0 \quad (2)$$

The conditions on $z$ are the regular strong concavity conditions for consumption goods. Time spent traveling is a "bad," which the individuals would be willing to pay to avoid. Concavity with respect to $t$ implies an increasing marginal disutility—the more time spent traveling, the greater the disutility from additional travel time.

The individual must allocate his total number of trips among the two modes:

$$x^i = x_1^i + x_2^i \quad (3)$$

The trip time by the highway mode is an increasing function of the number of trips taken by all travelers:

$$t_1 = d + a\left(\frac{X_1}{v - X_1}\right)^b \tag{4}$$

where

$X_1 = \sum_{i=1}^{I} x_1^i$, the total number of trips by all travelers via the highway mode;
$d$ = uncongested, "free-flow" travel time;
$v$ = capacity constraint of the highways (the upper bound on the number of trips that could be taken by highway, which would result in gridlock and an infinite trip time); and
$a, b$ = structural parameters reflecting the speed-volume relationship of the highway network.

The high-capacity transit mode is assumed to be completely unaffected by additional trips, and the trip time is a fixed value:

$$t_2 = c \tag{5}$$

The transit mode is assumed to be a high-speed mode, where the line-haul segment of a journey is rapid relative to, say, the expressway segment of a highway journey, thus compensating for slower speeds accessing the high-capacity mode including walk and wait times.

Equation 5 expresses the absence of an external effect from additional riders on the high-capacity mode. Of course, crowding on transit results in some riders standing and other inconveniences. However, the key operational assumption is that travel times on the high-speed mode are unaffected by changing volumes of passengers, which corresponds to the actual scheduling practice in rail transit systems.

Time spent commuting is given by the sum of trips weighted by the average time per trip. The $i$th commuter's total travel time is given by

$$t^i = x_1^i t_1 + x_2^i t_2 \tag{6}$$

The total trip time by the individual can be expressed as a function of the number of highway trips by substituting Equations 4 and 5 into Equation 6:

$$t^i(x_1^i) = x^i c + (d - c) + a\left(\frac{X_1}{v - X_1}\right)^b \cdot x_1^i \tag{7}$$

The first-order conditions of utility maximization are given by

$$P_1 - P_2 = \frac{u_{x_1}^i}{u_z^i} = \frac{u_t^i}{u_z^i} \cdot \frac{\partial t^i}{\partial x_1^i} \tag{8}$$

where

$$\frac{\partial t^i}{\partial x_1^i} = (d - c) + a\left(\frac{X_1}{v - X_1}\right)^b \cdot \left[1 + \frac{x_1^i \cdot b \cdot v}{(v - X_1) \cdot X_1}\right]$$
$$= t_1 - t_2 + \left(\frac{abv}{v - X_1}\right) \cdot \left(\frac{x_1^i}{X_1}\right) \cdot \left(\frac{X_1}{v - X_1}\right)^b \tag{9}$$

Some individuals will maximize utility by choosing all trips by one mode or another. However, some individuals will find their optimum allocation of trips by a mix of trips on both modes. These are "casual" switchers—that is, their circumstances or preferences do not lock them into a particular mode—and they correspond to the modal explorers discussed earlier. Equation 9 can be rearranged to give

$$\left[(P_1 - P_2) \cdot \frac{u_z^i}{u_t^i}\right] - \left[\left(\frac{abv}{v - X_1}\right)\left(\frac{x_1^i}{X_1}\right)\left(\frac{X_1}{v - X_1}\right)^b\right] =$$
$$t_1 - t_2 \tag{10}$$

or the condition under which door-to-door journey times across modes will be equal is given by

$$\left[(P_1 - P_2) \cdot \frac{u_z^i}{u_t^i}\right] = \left[\left(\frac{abv}{v - X_1}\right)\left(\frac{x_1^i}{X_1}\right)\left(\frac{X_1}{v - X_1}\right)^b\right] \tag{11}$$

Equation 11 indicates what combinations of prices, congestion, personal preferences, and highway speed-flow relationship will result in equal travel times. However, under the assumptions described earlier—especially the assumption of a growing marginal disutility with respect to travel time—it can readily be shown that with sufficient levels of congestion both the left-and right-hand sides of Equation 11 approach zero.

What happens under congested conditions? The left-hand side tends to zero because of the growing marginal disutility from increased travel time (also, the left-hand side approaches zero with increasing income—the individual becomes indifferent to the price differential as trip cost consumes a smaller portion of income). The theory also implies that congestion pricing will be less effective as congestion becomes more severe. It can be readily shown that if $u_t^i$ is not bounded, then for any combination of prices and capacity equation parameters and for any small value $\varepsilon > 0$, there is a level of congestion (number of total trips) sufficiently large such that

$$|t_1 - t_2| < \varepsilon \qquad (12)$$

## Empirical Evidence

Equations 10 and 11 tell us that if congestion is severe enough, journey times will tend to equal the journey time by the transit mode under the assumption of growing marginal disutility. This assumption can be tested empirically by estimating the relationships between travel time differentials, congestion, and additional factors.

### Source of Data

In an ongoing study for the Federal Transit Administration, door-to-door travel time tests were conducted on 14 urban corridors. The testing was conducted between February and June 1995. The corridors were selected on the basis of criteria that included congestion, population density, the existence of mature dedicated-guideway transit systems, and public transportation headways. The 14 corridors where data was collected are given in Table 2. The corridors span a range of moderate to high congestion. In each corridor random routes of origins and destinations were selected. Survey crews conducted peak-hour trips on the different modes under comparable conditions.

More than 1,000 trips were recorded, and some of the average results are reported in Table 1. Of the trips taken, 495 pairs of comparable automobile/transit trips were observed. Congestion data for the metropolitan areas in which each of the corridors was located were taken from the recent TRB study on urban congestion (*4*). The metropolitan planning organizations in each corridor provided information on transit headways.

### Analysis of Data

A regression analysis of time differentials was conducted. The absolute value of the travel time difference, automobile versus transit, was regressed against the metropolitan area congestion index and the transit mode headway (minutes). The results are presented in Table 3. The two explanatory factors, congestion and headway, do little to explain the variation between each of the 495 trip pairs. This is not surprising, since these variables have no variation within the corridor and transit mode. However, we observe that the coefficient for congestion is negative whereas that of headway is positive, and both coefficients are significant at the 99 percent level. This means that travel time differentials diminish with growing congestion and increase as transit headways increase.

Undoubtedly there are additional factors that contribute to the explanation of travel time differentials, some of them location specific and others associated with price and other variables. However, we find that the evi-

TABLE 2 Corridors Studied

| Corridor | Modes |
|---|---|
| Boston - Mass Pike | Auto, Commuter Rail |
| Boston - Southeast Expressway | Auto, Heavy Rail |
| Chicago - Midway | Auto, Heavy Rail |
| Chicago - O'Hare | Auto, Heavy Rail, Commuter Rail |
| Cleveland - Brook Park | Auto, Heavy Rail |
| Philadelphia Schuylkill - Bryn Mawr | Auto, Commuter Rail |
| Philadelphia Schuylkill - Upper Merion | Auto, Commuter Rail |
| Philadelphia - Wilmington | Auto, Commuter Rail |
| Pittsburgh - Parkway East | Auto, Express Bus |
| Princeton - New York | Auto, Commuter Rail |
| San Francisco - Bay Bridge | Auto, Commuter Rail |
| San Francisco - Geary | Auto, Express Bus |
| Washington - I-66 | Auto, Heavy Rail, HOV |
| Washington - I-270 | Auto, Heavy Rail |

TABLE 3 Regression Results

| *Dependent Variable: Absolute Value of Trip Time Difference (Auto - Transit)* | |
|---|---|
| **Variable** | **Coefficient (t-values)** |
| Constant | 21.51 ( 5.54) |
| Congestion Index | -4.743 (-2.61) |
| Headway | 0.2703 (4.07) |
| *All coefficients are significant at the one percent level* | |
| **Summary Statistics** | |
| Number of Observations | 495 |
| $R^2$ | 0.051 |
| Mean Dependent Variable | 15.63 |
| F-Statistic | 13.18 |

dence supports the theory that in congested urban corridors the growing marginal disutility from time spent traveling causes door-to-door journey times to converge to the journey time by the high-capacity transit mode. Furthermore, the data indicate that reducing transit headways (which, in general, will contribute to shorter trip times by transit) will also contribute to a reduction in the time differentials between modes.

## Implications for the Benefit-Cost Analysis of Transit Investments

The preceding analysis indicates that the observation of equal or nearly equal travel times across modes is consistent with consumer theory and may be observed under a wide range of circumstances with high levels of congestion. Congestion, if severe enough, will drive a multimodal transportation system toward convergent travel times. The further empirical study of congested corridors will reveal which combination of underlying factors (economic, demographic, spatial-locational, etc.) are most closely associated with the condition of travel time convergence. Travel time convergence in congested urban corridors and the factors promoting that convergence should be crucial elements in the development of transportation policies, especially in an environment of budgetary constraint with congestion pricing a rarity.

The benefit-cost analysis of transit investment examines the demand for trips and derives consumer surplus estimates based on the schedule of demand. The non-transit trips are mostly assigned to the highway network, and cost savings from reduced congestion are estimated. Trips are allocated between modes using a modal choice algorithm that does not take into account the dynamic interaction between the modes. When the allocated trips are assigned to the highway network, even under highly congested conditions, forecast journey times will likely be highly divergent.

As a first step toward refining the benefit-cost analysis of transit investment with a view to accounting for the phenomenon of convergence in congested corridors, the analyst should examine whether the modal split will yield journey times consistent with the convergence dynamic after trips are assigned to the urban transportation network. If convergence is likely to occur in the corridor under analysis, there is strong theoretical and empirical justification for calibrating the modal constants in the modal choice model such that the assignment of traffic yields nearly equal journey times.

### ACKNOWLEDGMENT

The author thanks David Lewis and Martin Mogridge for their motivating thinking on multimodal performance and benefit-cost, Herbert Mohring for his helpful suggestions in developing the microeconomic theory, and Steven Lewis-Workman for his thoughtful review of an early draft. Support for the research by the Federal Transit Administration is gratefully acknowledged. Special thanks are owed to Fred Williams of the Federal Transit Administration's Office of Policy for his efforts in facilitating the research.

## REFERENCES

1. Mogridge, M. J. H. *Travel in Towns: Jam Yesterday, Jam Today and Jam Tomorrow?* The MacMillan Press, Limited, London and Basingstoke, England, 1990.
2. *Unsticking Traffic: When Transit Works and Why.* Federal Transit Administration Policy Paper, Oct. 1994.
3. Downs, A. *Stuck in Traffic: Coping with Peak-Hour Congestion.* The Brookings Institution, Washington, D.C., 1992.
4. *Special Report 242: Curbing Gridlock: Peak-Period Fees To Relieve Traffic Congestion.* TRB, National Research Council, Washington, D.C., 1994.