APPENDIX B

INDUCED TRAFFIC AND INDUCED DEMAND

Douglass B. Lee, Jr.

“Induced” is a term implying that a particular condition is indirectly caused by another condition. In the case of traffic volumes, the term arose from the phenomenon that improvements to a highway -- especially capacity improvements -- seemed to result in more traffic choosing to use the road than would be the case if the highway were not improved. To an economist, this is an example of demand elasticity. Simply recognizing that travel demand is elastic, however, is not sufficient to reconcile the conflicting views of engineers, planners, and environmentalists. On one side are those who argue that transportation facilities are provided to serve land uses and support economic activity; on the other are those who claim that whatever capacity is provided soon fills up to the same level of congestion, gaining nothing. The truth can be better understood by defining induced demand in a way that uses the concept of elasticity.

This appendix describes the concepts guiding several modifications that were made to the HERS model for the 1997 Conditions and Performance report to Congress. With minor exceptions noted below, the model implements the concepts as they are described here.

Concepts of Induced Demand

Frequent references are made in transportation planning to the concept of induced demand, but the term remains ambiguous. The intent here is to define the relevant concepts, and show how they can be operationalized in representing demand for purposes of benefit-cost evaluation of capital improvement projects.

Acknowledgments: The author thanks Ross Crichton, William Goldsmith, and Anthony Rufolo for valuable comments and suggestions. Lisa Klein and Gregorio Camus were instrumental in the development of the algorithms.
Exogenous Demand Factors

Historically, demand forecasts in urban transportation planning have been based on exogenous variables such as land use, population, employment, and income. Once these variables are measured or estimated, the result is a “point” estimate for traffic volume at a future date. Demand, in this sense, is influenced by neither transportation infrastructure nor money price, but is determined entirely by exogenous factors.

If demand is determined by forces beyond the control of the transportation planner, then failure amounts to not having adequate facilities to handle it, and the planner is simply the messenger. Alternatively, if the facility creates its own demand, the planner is just furthering the careers of planners.

Demand Fills Capacity

A contrasting concept has emerged claiming that additional capacity stimulates corresponding increases in demand. This concept embodies the “build it and they will come” idea, or a belief in the existence of “latent demand,” which suggests that there are willing buyers who will express their demand for travel once the service is offered.\(^1\) In growing urban areas, the evidence from recent decades seemed to support this interpretation.

Although the idea has not been implemented as a formal forecasting method, the implication is that demand is entirely endogenous. If true, the policy choice is whether to permit travel to grow or to suppress it.

Elastic Demand

Perhaps the first recognition that demand responded to endogenous factors was the assertion that congestion is self-regulating, implying an automatic balancing of supply and demand. More recently, the economist’s concept of demand being a relationship between price and quantity demanded has become accepted, if not necessarily applied in practice. From this perspective, all endogenous changes in volume are movements along the demand curve, whether they are called latent, induced, or something else. If “price” is generalized to include travel time, operating costs, and accidents, then changes in capacity and alignment alter the “price” and thereby cause movements along the demand curve.

Overall, then, travel demand is the result of a combination of both exogenous factors that determine the location of the demand curve, and endogenous factors that determine the price-volume point along the demand curve.

Short Run versus Long Run

The short run can be any period of time over which something remains fixed. What is fixed might be the capacity of a highway, fuel efficiency of the vehicle fleet, locations of

---

\(^1\) For an interpretation of latent demand, see Small (1992), pp. 112-116, or Small, Winston, and Evans (1989)
short run versus long run

Appendix B

Induced Traffic and Induced Demand

Short Run versus Long Run

Employment, or anything else that changes slowly. The long run is enough time for these characteristics to change. The short run is typically assumed to be about a year in transportation planning, but the dividing line depends upon the practical context.

Demand “elasticity” is the responsiveness of quantity demanded to changes in price. Price is generalized for travel demand purposes to include travel time, operating costs, and accidents, as well as user charges. Everything included in this generalized price is an endogenous factor with respect to induced traffic. An increase in capacity that lowers travel time, for example, results in additional travel if the elasticity is not zero.

Short run demand elasticity tends to be lower (less elastic) than long run elasticity, because more opportunities to increase or reduce consumption can be developed over the long run than in the short run, while short run options do not diminish in the long run. If the price of fuel goes up, for example, highway travelers can reduce fuel consumption by taking fewer trips and chaining trips together, by carpooling to share expenses, by driving in ways that achieve better mileage, and by taking a larger share of trips on transit. In the long run they can also switch to more fuel-efficient vehicles, and change their workplace and residence locations. If the price stays high, vehicle manufacturers will develop and produce more fuel-efficient vehicles, and better transit service may be offered.

While the distinction between short run and long run demand is really a continuum rather than two discrete states, the separation is useful both conceptually and for modeling purposes. In Figure B-1, two short run demand curves are shown in relation to their common long run demand curve (the latter indicated by a dashed line). Demand could be for a facility, a corridor, or even travel in a region. At a “long run” price of $p_1$ the volume is $v_1$ and the short run demand curve $D_1$ applies, such that changes in the price cause changes in volume along this demand curve in the short run. If the price drops to $p_2$, for example, then volume will increase to a flow of $v_{1,s}$. If the price stays at that level for the long run, then the short run demand curve will shift outward to $D_2$, resulting in the volume $v_2$ at that price. If the price were then to go back up to $p_1$, volume would only drop to $v_{2,s}$ in the short run, but eventually back to $v_1$ in the long run.

For example, secular declines in real fuel prices have led to increases in the size and weight of vehicles and concomitant declines in their fuel economy; if the price of fuel were to increase, gasoline consumption would drop but the vehicle fleet would take time to evolve to a more fuel-efficient average. Changes are not necessarily completely reversible: knowledge gained from research leading to advances in technology in, say, fuel efficiency, is not lost when the need is lessened, but its application tends to diminish.

2 The generalized price embodied in HERS includes time, operating costs, and accidents, but no user charges per se. The implications of this omission are discussed in greater depth in Appendix D.
A similar distinction can be made between “induced traffic” (or induced travel) and “induced demand,” by applying the short run and long run concepts: demand is assumed fixed in the short run, so changes in volumes are the result of movements along the demand curve, whereas the short run demand curve can shift in the long run. Thus these terms are defined such that “induced traffic” is a movement along the short run demand curve, while “induced demand” is a movement along the long run demand curve, or an endogenous shift in the short run demand curve.

In Figure B-1, no time direction is implied on the horizontal dimension; the shape of the long run demand curve does not mean that price declines over time. Nor are the short run demand curves necessarily ordered from one to two; demand could start at $D_2$ and then shift to $D_1$. The diagram shows only the relationship between price and volume under short run and long run conditions.

Long run elasticity -- as with any other demand elasticity -- is a ratio of the percent change in quantity demanded to the percent change in the price of the good. Referring to Figure B-1, the first circled point at $(p_1,v_1)$ is taken to represent a point on both the short run and long run demand curves. The second circled point at $(p_2,v_2)$ represents the long run result of a price change, which lies on the previous long run demand curve but a new short run curve. The arc elasticity between the two points is

**FIGURE B-1.** Long run demand with short run demand curves.
where $e_{LR}$ is the long run elasticity of demand. If the following simplifications are made for ease of presentation,

$$
a = p_2 - p_1
$$

$$
b = v_{1,s} - v_1
$$

$$
c = v_2 - v_{1,s}
$$

as shown in Figure B-1, then the long run elasticity can be represented as

$$
e_{LR} = \frac{b + c}{a} \times \frac{p_1}{v_1} = \left( \frac{b}{a} \times \frac{p_1}{v_1} \right) + \left( \frac{c}{a} \times \frac{p_1}{v_1} \right)
$$

where the first term in parentheses is the short run elasticity ($e_{SR}$) and the second term is the shift in the demand curve over the long run, represented as an elasticity. Thus the long run elasticity is the sum of the $e_{SR}$ and a purely long run component which will be called the long run share, $e_{LRS}$, defined as

$$
e_{LRS} = \left( \frac{c}{a} \times \frac{p_1}{v_1} \right) = \left( \frac{v_2 - v_{1,s}}{p_2 - p_1} \right) \times \frac{p_1}{v_1}
$$

so

$$
e_{LR} = e_{SR} + e_{LRS}
$$

The $e_{LRS}$ component can be interpreted in the same way as a normal elasticity, and can be empirically measured as the difference between the short run elasticity and the long run elasticity estimated for the appropriate time period.$^3$

**Induced Traffic**

As defined above, induced traffic is a movement along the short run demand curve. Common usage of the term “induced” suggests additional traffic, i.e., an increase in volume. Decreases might be called disinduced or deterred or discouraged traffic. For present purposes, the term induced refers to any endogenous change, whether positive.

---

$^3$ See Taplin (1982) for theory.
or negative. Increased congestion or higher tolls, other things being equal, will cause a reduction in volumes. If this occurs in the short run, this is (negative) induced traffic.

Some of the possible sources of induced traffic are:

- Diverted traffic that changes its route onto the improved facility.
- Rescheduled traffic that previously used the facility at a different time (spreading or contracting the peak).
- Shifts from other modes -- which may or may not have used the facility before -- including changes in occupancy.
- Destination shifts resulting from the improvement of the facility.
- Additional travel by persons already using, or in the market for, the facility.

Demand forecasts for a new or improved facility always include at least some of these sources, although such estimates seldom explicitly recognize a generalized price as the explanatory variable and do not produce a schedule of price-volume combinations.

**Partial and General Equilibrium Demand Curves**

All demand curves portrayed in this analysis are assumed to be general equilibrium demand curves, even those for the short run. Thus they include traffic shifted to or from other modes or from alternative facilities. A partial equilibrium demand curve, as repre-
Presented in Figure B-2, includes only the travel for those already in the market, whether they are currently taking trips or not (e.g., a person who did not travel at all in this corridor but who chose to do so after the price was reduced, and not by shifting a trip from another time or place). If the demand curve includes diverted travelers (from other modes, routes, times, or destinations), then it will be more elastic than the corresponding partial demand curve because more options are offered. Thus some of the (short run) induced travel comes from new trips by persons already in the market, and some comes from trips diverted from other markets.

For every point on the general equilibrium demand curve there is a corresponding partial demand curve, representing the (hypothetical) demand that would occur if there were no substitution between markets. If the price were raised, for example, from a point on the general equilibrium demand curve, a movement up the partial demand curve would imply that the travelers could not divert to another time or facility. Not surprisingly, such a demand curve cannot be observed in practice.

Because demand forecasts usually include diverted trips, practical demand forecasts are aimed implicitly at constructing (or locating points on) a general equilibrium demand curve. If the demand is for a single facility, then induced traffic will appear large relative to previous volumes, because most of the change in trips will be diversions. At the regional level, induced traffic -- if it were actually estimated -- would be a smaller share of total traffic growth because only trips diverted from other regions, plus substitutions between transportation and other goods, make up the induced share. For project evaluation, diverted travel and other components of induced demand as measured in consumer surplus represent the net valuation of systemwide impacts.

In Figure B-2, all of the movement along the general equilibrium demand curve stimulated by the reduction in price from $p_0$ to $p_1$ is labeled “induced trips.” A portion of this induced traffic is labeled “diverted trips.” If the diverted trips are removed from the total “gross” induced traffic, the residual might be called “net” induced traffic. Some analysts prefer that the term induced be restricted to mean net induced trips, and the others be left as diverted trips.4

For some purposes, this usage has an appeal, but the distinction is a difficult one to make. A trip between the same origin and destination but using a different route is clearly a diverted trip, but trips at other times, or to other destinations are less obvious. If the improved facility prompts me to go to a movie instead of renting a video, and the video store is much closer, is this induced or diverted? Suppose I would have walked to the video store? Suppose I would have had the video delivered, and the van would have used the same facility before it was improved? What can be observed directly is that more vehicles use the facility after it is improved, and that trips in the region do not go up by as large an amount as the volume on the improved facility. Labeling which particular travel is “new” and which is “diverted,” however, is difficult and probably not necessary.

---

4 Examples include Dowling (1994) and SACTRA (1994).
Schedule Delay and Peak Shifting

As noted above, changes in the generalized price may lead to changes in schedule. Peak congestion may be at least partially avoided by leaving earlier or later than preferred. A reduction in peak travel time will cause some travelers to join the peak because the cost to them of schedule delay (departing at a different time than preferred) is less than the new peak delay. Thus induced traffic may be diverted from other times as well as other routes.

If the demand curve represents both peak and off-peak, then the elasticity will be lower than if peak is separated from off-peak. Because the two periods are so closely interrelated (off-peak demand depends upon peak price, and vice versa), separating them for benefit-cost purposes can be tricky, but that is one way to include benefits from reducing schedule delay.

**Induced Demand**

For purposes of evaluating costs and benefits, the overall analysis period for a project (generally the project lifetime, e.g., twenty years) is broken into a series of discrete time

---

periods, during each of which the demand curve is assumed to be fixed. A baseline long range forecast is used to establish the short run demand curve for each period.

A demand forecast is a functional relationship between time and traffic volume, assuming a set of conditions. Exogenous conditions include population growth, economic growth, land use patterns, and available substitute transportation alternatives. Endogenous conditions include capacity, level of service (LOS), and user fees. For the present analysis, all endogenous factors are represented in the generalized price. Capacity and LOS, for example, would both be subsumed under travel time cost, and included in the generalized price.

The baseline long run demand forecast assumes a generalized price, as well as whatever exogenous factors are thought to be relevant by the forecaster. Alternative forecasts under different assumptions might be constructed, as shown in Figure B-3. One such forecast is selected for constructing the short run demand curves.

The distinction between long run induced demand and short run induced travel is implemented by constructing a short-run demand curve for each of the shorter demand periods (e.g. 1-5 years), and allowing the initial curve to shift depending upon previous improvements. The forecast becomes a series of discrete points, shown circled in Figure B-4, that provide the calibration points for the associated short run demand curves. The short run demand curve can be a straight line calibrated with an elasticity, or a constant elasticity demand curve, or some other functional form that can be fitted to a single price-quantity combination. The elasticity chosen should be appropriate to the length of the demand period.\(^7\)
A single fitted short run demand curve is shown in Figure B-5, along with other relevant prices and volumes. The price from the previous period \( p_{\text{final, } t-1} \) is adjusted to account for traffic growth, pavement wear, accident rates, and user fee changes that have occurred since the previous period. The result is \( p_{\text{no improvement}} \). Alternative improvements for the current period are evaluated, and, if any are feasible, the best is implemented. This results in the \( p_{\text{improved}} \) price, which becomes the initial price for the next demand period. If no improvement is selected, the unimproved price carries into the next period.

Evolution of demand in the long run is built upon what takes place in the short run. Operationally, induced demand is defined to be the shift in the short run demand curve caused by the price in the previous period. If the price in all previous periods is the same as the baseline price, then the demand curve is fitted to the baseline forecast for that period. If an improvement is made in one period that reduces the price below the baseline price, this leads to a shifting of the demand curve outward, according to the percent by which the price in the previous period is below the baseline price. If no improvement is made, the price increases relative to the baseline forecast price, and the demand curve shifts inward in the next period. These two possibilities are shown in Figure B-6. For example, a price of \( p_{\text{no improvement}} \) will shift the subsequent demand curve inward from \( q_{\text{forecast}} \) by a percentage equal to \( (p_{\text{baseline}} - p_{\text{no improvement}}) \times e_{\text{LRS}} \).

Currently, the demand period or “funding period” in HERS is five years, so the short run elasticity should be selected to allow for adjustments that can be expected to take place within that span of time.
The relationship between the difference in price between the final (improved or not improved) price and the baseline price, for one period, and the horizontal shift in the demand curve in the next period, is governed by the long run share $\epsilon_{LRS}$, as described above. There is no long run demand curve as such, but the shift attributed to induced demand is a displacement of the short run demand calibration point along the baseline price line.

Incorporating induced demand, then, allows each period's demand curve to be a function of the previous period's investment (as it affects price to the user). Investment that keeps the price in each period below the baseline price for the baseline forecast produces demand curves that shift outward farther and farther, compared to the baseline forecast. Similarly, if improvements are not made and price is allowed to rise in each period (due to congestion, pavement roughness, and accidents), the demand curve will be continuously shifted inward relative to the baseline.

The magnitude of this shifting -- the sensitivity of long run demand to investment and pricing -- is determined by the $\epsilon_{LRS}$ parameter. The shorter the time period for the short run, the lower should be the long run elasticity shift from period to period. If the long run induced demand parameter is zero, the location of each short run demand curve would be determined by the baseline forecast, without regard for which, if any, improvements were made in any demand period. Short run movements along the demand curve could still occur, depending upon the short run price elasticity, but there would be no cumulative endogenous effects from one period to the next. Alternatively, with a high

---

7 See “Disaggregation of Long Run Elasticity” on page B-4.
$e_{LRS}$, induced demand could alter the baseline forecast, even to the point of potentially offsetting the trend of the initial forecast, such leading to growth in demand (from keeping the price low) despite a declining forecast, or causing a decline in demand despite a growth forecast (traffic is deterred by congestion and bad pavement, as a consequence of no improvements).

**Getting to the Long Run**

Empirical estimates of the two elasticities depend upon the length of the short-run time period and the rate of adjustment to changes in price. The length of time between a change in conditions and a new equilibrium is somewhat arbitrary because other conditions change before equilibrium is reached, but the process is one of accelerating initial response followed by gradual refinement. In the context of highway volume adjustments in response to changes in the generalized price of travel, the short run is up to a year. The long run -- allowing for changes in residence and workplace locations -- begins within a year but may not run its course for upwards of twenty years. Such changes are not likely to be motivated solely by changes in transportation prices, but may take transportation user costs into account when the change is made for other reasons (new job, change in income, change in family).

An approximate adjustment curve is shown in Figure B-7. Although the curve is not fitted to specific data, it reflects the generally observed pattern that roughly half the adjustments take place within about a quarter of the time to long run equilibrium. If the full long run adjustment period is 10-20 years, then half the long run elasticity occurs within the first 2.5 to 5 years. There might be some accelerating adjustment in the first year, as

---

shown, based on the idea that responses don’t occur until consumers become sure the price change will stick, or begin feeling its effects.

Many studies have estimated travel demand elasticities, but one of the difficulties in interpreting these results is the uncertainty of the time frame that is applicable to the data. Another confounding problem is the ambiguity of the base of the observed elasticity; because most of the empirical cases observe a change in a small component of the total price of travel, the base for computing the percentage change in price is often not obvious and may not be given explicit treatment. The potential differences are large, e.g., a factor of three or more.9

The parameter sought is the elasticity of vehicle travel with respect to its own price, including user fees, operating costs, and travel time. Studies undertaken to date suggest that short run elasticities tend to fall in a -0.5 to -1.0 range, and long run elasticities from -1.0 to -2.0; a within-period short run elasticity for a 5-year period would thus be -0.6 to -1.0 and the between-period elasticity from -1.0 to -1.6, yielding an $\epsilon_{LRS}$ of about -0.4 to -1.0.

Two aspects of the demand forecast are of particular interest. One is how to impute a presumed price to the baseline forecast. The second is whether long run feedback of transportation investments on the demand curve has been incorporated into the forecast.

(1) Baseline Price. Although the generalized price behind a demand forecast is seldom made explicit, such attributes as LOS and accident rates may be, and others can be guessed. Pavement quality is probably assumed to be good, and operating costs are typical for the conditions (terrain, vehicle type, congestion). As a default, the current LOS can be assumed.

(2) Long Run Demand Feedback. Constructing or expanding a facility stimulate or permit some additional travel in the long run even if the price is unchanged from the baseline. Hence, the baseline forecast should include growth in travel that will result from traffic-generating activities that choose to locate in such a way as to take advantage of the services provided by the facility, at the baseline price. The long run elasticity amplifies this effect up or down, but does not substitute for it.

If forecasts are based on historical patterns over a time horizon of half a dozen years or more, then the feedback effect is implicitly built in. Whether it needs to be made explicit or refined is an open question, but the impacts of errors in out-year forecasts are suppressed somewhat by discounting.

[9] The empirical evidence and methods for estimating highway travel demand elasticities are covered in Appendix C.
Summary

Some of the ambiguity and confusion that surrounds the discussion of induced demand might be dispelled by applying the following definitions and principles:

(1) The term *induced* means a movement along a travel demand curve as a result of changes in *endogenous* factors, which can be represented as components (time, running cost, money) of a generalized price.

(2) The measurement of induced travel is dependent upon the *market* for which the demand curve is defined; induced travel defined at the facility level will include traffic diverted from parallel routes, while induced travel at the regional level will include only trips that are new to the region.

(3) A useful distinction can be made between short run demand and long run demand: movements along the *short* run demand curve amount to *induced traffic*, whereas movement along the long run demand curve constitutes a *shift* in the short run demand and can be called *induced demand*.

(4) Benefit-cost evaluation of projects requires that baseline demand forecasts be adjusted to take into account induced demand, both short and long run; this is simply to say that improvements that change user costs should be evaluated in the light of whatever changes in volume will actually occur. Such demand curves are referred to as general equilibrium demand curves.

(5) If the short run elasticity is zero, then traffic volumes are unresponsive to changes in price within a single demand period, and the demand curve is vertical. If the long run share (i.e., excluding short run effects) elasticity is zero, then there are no long run effects (e.g., no investment in highway-related facilities or land use changes) stimulated by highway pricing and investment policies. Empirically, neither of these conditions seems to apply.

References


