Induced Demand:

Its Definition, Measurement, and Significance

Don Pickrell
Chief Economist
Volpe Center, U.S. Department of Transportation

Prepared for:
Eno Transportation Foundation Policy Forum
February 22-23, 2001
Washington, D.C.

Don Pickrell
Draft 3/27/2001

During the heyday of highway building in the United States, engineers and planners were frequently astonished by the discovery that newly-opened highways quickly filled to near their design capacity. There are many stories illustrating this point, but one repeated frequently was that Los Angeles' Hollywood Freeway reached its "design-year" capacity – a level of usage that was supposed to require twenty years to approach – within two years of its completion. By the time I joined the profession in the 1970s, this phenomenon had come to be widely known as "induced travel," an acknowledgement that the increase in travel observed on new highway segments was somehow caused indirectly by their construction. Still, the process by which it occurred seemed to mystify transportation planners and highway engineers, who were generally unable to explain its origins.

The phenomenon of induced travel was so difficult for planners and engineers to understand because the forecasts of total travel demand they prepared for new facilities were based entirely on exogenously-determined factors such as the spatial distributions of population and employment, together with demographic characteristics of households. Forecasts of the geographic distribution of trip destinations, automobile mode shares, and drivers' choices of specific routes often recognized the roles of travel time and costs, but projected levels of total travel demand almost never did. Total demand forecasts were simply point estimates of future travel (usually vehicle) volumes, which were determined entirely by these exogenous factors, influenced by neither the travel speeds permitted by the highway network nor costs borne by motorists in using it.

Downs' "Iron Law" of Traffic

Highway engineers and planners clearly understood that travel speeds were influenced by the volume of vehicles using a facility, but probably the earliest recognition that travel demand varied in response to speed came as part of Downs' (1962) realization that congestion is likely to be self-regulating. He argued that a slightly different version of the normal balancing between demand

and supply through the mechanism of price that operates in economic markets also operated on highways. There, the sensitivity of demand to travel speed and the speed-reducing effect of congestion – which produces the usual upward-sloping supply curve found in economics textbooks -- would interact to balance demand and road capacity, but often at a traffic volume that produced considerable congestion.[1]

As a consequence, expanding the capacity of a highway would increase travel speeds, thereby causing increased demand for its use, and a new equilibrium travel volume and speed would ultimately be established. Downs argued that the primary sources of this added demand were likely to be diversion of travelers to the improved facility from parallel routes, competing travel modes, or other hours of the day. Depending on the sensitivity of demand to speed, the new volume could be well above the pre-improvement demand level, in which case the new equilibrium speed might actually be closer to its original level than to the free-flowing speed forecast for the expanded facility.

"Capacity-Induced" Travel Demand

More recently, the critical role of speed -- the dominant component of price for many transportation services, particularly those provided by highways -- in equilibrating demand and capacity has become increasingly overlooked by the emerging view that investments in additional transportation capacity stimulate corresponding increases in demand. This view emphasizes the existence of a direct linkage between expansions in the capacity of highways or other transportation infrastructure and their level of usage, with the exact mechanism that establishes this linkage, namely the effect of increased capacity on travel speed, left implicit or overlooked completely. It stresses the importance of "latent de-

---

[1] Much later, Morris (1980) explicitly argued that traffic levels were determined by the interaction between the dependence of travel demand on speed and the upward-sloping "supply" curves – or relationships between travel speed and traffic volumes -- that characterizes roads and highways.

mand," or the existence of willing travelers who will demonstrate their demand to use a transportation facility -- a highway, airport, or ?? - only after it is improved.

According to this view, this "induced demand" may be sufficient to restore congestion and slow travel speeds on an improved facility to levels near those before it was improved, often within a surprisingly short time after the improvement is completed. Advocates of this view seemed astonished to discover that it could operate in reverse, causing what was somewhat awkwardly referred to as "disappeared" or "suppressed" traffic in response to reductions in the effective capacity or outright closures of links in a highway network.

Travel as an Economic Good

My perspective is that the economic concept of demand as a relationship between the price of a service and the quantity of it that individuals and the market collectively demand applies to highways and other transportation facilities (or more formally, to the services they produce, such as the movement of people or vehicles). In the context of transportation services, price must be generalized to include travel time and costs incurred by travelers such as those for operating vehicles, as well as the usual money costs. Viewed from this perspective, changes in travel volumes in response to variation in this "generalized price" for using a transportation facility represent movements along a conventional economic demand curve, whatever terminology is applied to them.

More precisely, households and businesses demand the services produced by transportation infrastructure as an input to their processes of producing transportation services for their own use (in the case of households and some businesses) or for sale (as for firms whose business is providing transportation services for sale). Demand for these services is determined by a combination of demographic and macroeconomic factors, which establish the demand curve itself, and partly by their generalized price. The value of the generalized price locates a specific price-volume combination along the demand curve, and movements along this curve in response to changes in the time or user cost components of generalized price represent induced changes in demand for the services provided by the facility or network.

Defining Induced Demand

Figure 1 illustrates induced travel in the simplest possible terms, using a highway segment as an example. Its horizontal axis represents the quantity of services produced by the highway during some standard time period such as the morning peak travel period or over an entire day. This flow of services is measured by the equivalent number of standard-size vehicles (passenger cars are the usual *numeraire*) using the highway during that period; in other words, the hourly or daily volume of vehicles it carries. The vertical axis of Figure 1 illustrates the generalized price faced by each vehicle using the highway during the relevant time period, which consists of the value of its occupants' travel time, the costs of operating the vehicle over the segment,

The curve labeled D in Figure 1 is the demand function for travel on that highway segment during the time period in question. It relates the number of vehicles using it during that period to the generalized price per vehicle trip on the segment, and shows that more trips will be made as this price declines. In this respect, the demand curve for the transportation services produced by this (or any other) highway segment looks exactly like the demand curve for any other commodity or service. If the generalized price declines from an initial value of P_1 to a lower value P_2, the volume of trips made on the segment during the peak period or day will increase from V_1 to V_2. This increase in travel volume can be said to have been *induced* by the decline in the generalized price of making a trip. This response is completely symmetrical, so that if the generalized price subsequently returns to its original value P_1, it will likewise induce a decline in the volume of travel from V_2 back to V_1.

Capacity Expansion and Induced Travel

What might cause the generalized price of using the facility to decline? The largest component of the generalized price of a vehicle trip is typically the value its occupants attach to the time required to complete the trip in the particular circumstances – comfort, privacy, security, and so on – provided by the vehicle in which they are traveling. When an increasing number of vehicles attempt to use the highway facility at the same time, they interfere with one another in the process known as congestion. Strictly speaking, their drivers travel at progressively slower speeds as traffic becomes more dense and the spacing between vehicles -- following distances, in the parlance of driver training -- become shorter.

As a consequence of this behavior by drivers, the relationship between the travel time component of the generalized price per vehicle and travel volume on the faculty is upward sloping. Other components of the generalized price, such as vehicle operating costs, may also increase with growing travel volumes per time period, although probably none as rapidly as travel time. When the volume of vehicles traveling on the facility approaches the maximum it can carry, this "price function" slopes upward with increasing steepness.[2] Although it is tempting to think of this price function as the supply curve for the services produced by the highway facility, this analogy is not exactly correct for reasons that will be discussed in detail later in this paper.

---

[2] Reference is frequently made to the "design capacity" of a facility as if it was a fixed value, but in practice it is also determined partly by driver behavior. This occurs because the volume of vehicles moving through the facility is a product of their density (measured for example in vehicles per lane-mile) and the speed at which they are traveling. Because drivers travel more slowly as density gradually rises, their product attains a maximum value where the contribution to increased volume from higher vehicle density is exactly offset by the resulting slowing of travel speed, and then begins to decline. The physical design of a facility affects this maximum value by determining the rate at which driving speeds decline as the density of vehicles or traffic on the facility increases. A facility's maximum volume typically occurs at vehicle densities far short of the those it could accommodate if it functioned like a parking lot (its "jam density"), although this is exactly what many highways do at certain hours. This "capacity-reducing" effect, sometimes referred to as "hypercongestion," is a completely wasteful way for a transportation facility to operate.

In any case, both the volume of travel on the facility and the actual value of generalized price are determined by the interaction of the demand function for the facility and this price function. Figure 2 illustrates this "equilibration process" graphically, with the initial generalized price P_1 and travel volume V_1 both simultaneously determined by the intersection of the facility's demand curve D and its initial price function PF_1. Investments in expanding the capacity of the facility are designed to reduce the rate at which the generalized price faced by its users increases with growing travel volumes; or said another way, to allow it to accommodate higher travel speeds at all volumes. Investments of this type can take different forms – adding lanes is probably the simplest and most familiar – and they may expand the theoretical maximum vehicle-carrying capacity of the facility, but their more important consequence is to slow the increase in generalized price from the value associated with free-flowing travel speeds as usage of the facility grows through the range it normally experiences.

Because faster travel speeds lower the generalized price, this is be reflected a downward shift in the price function for the facility, to something like the position labeled PF_2 in Figure 2. In turn, the effect of this downward shift in the price function is to establish a new equilibrium with the demand function for the improved facility, which as Figure 2 illustrates occurs at a lower generalized price P_2 and larger travel volume V_2. The usual -- but somewhat imprecise -- interpretation is that the increase in volume (equal to V_2 minus V_1) has been *induced* by the investment in expanding the capacity of the facility, when in fact it has been induced by the decline in the generalized price faced by users, itself a consequence of the downward shift in the facility's price function brought about by the investment. Although this sounds like a needlessly pedantic distinction, its importance will become clear shortly, and it also corresponds more closely to the English-language meaning of the term induced, which my dictionary defines as "brought about through an indirect influence."

Induced Travel in the Short and Long Run

The short run is usually defined in economics as the period of time over which the level of capital investment in production facilities remains fixed, but other factors can be varied. While this isn't an entirely helpful definition from the standpoint of travel demand, probably the closest analogy is the period during which households' residential locations as well as the spatial distribution of economic activity – and thus of employment -- remain fixed. Although in practice some households and businesses are always in the process of relocating, we can think of the short run as the period during which a typical household's residential location, as well as the employment status and workplace locations of its members, remain unchanged. In contrast, the long run is a period sufficiently long for one or more of these factors to change, again for a typical household.

In the short run, the demand for travel is expected to be less sensitive to changes in generalized prices on an area's highway network, since household members' commitments to participate in activities away from home may not be easily rescheduled or renegotiated. Thus its opportunities to economize on travel demand are likely to be limited to linking individual trips into "chains," altering usual travel routes, or changing modes of travel. Over the longer run, however, travelers can reorganize the number and sequence of outside-the-home activities that generate their demands to travel, change the locations at which they participate in some of these activities, or even relocate their homes or jobs. Households can also adjust the number and specific types of vehicles they own -- probably much more quickly than they can move or change jobs, in fact – as ways of modifying their demands for the services of highways. These responses allow travel demand to be more sensitive to variation in the pattern of generalized prices for travel on the network in the long run than in the short term.

Graphically, the greater sensitivity of demand to travel conditions in the long run will be shown by the short-run demand curve being steeper than its long-run counterpart, which will appear "flatter." Figure 3 shows two short-run demand curves, labeled D_1 and D_2, in relation to their common long-run demand curve, which is indicated by D_{LR}. Each of these curves could illustrate demand for the use of a specific transportation facility, for travel within a single corridor, or even travel over an entire network. At an initial price of P_1 the volume of facility use or travel is V_1, and changes in travel speed on the facility or dollar costs incurred by its users will cause movements along the demand curve D_1 in the short run. For example, if the generalized price for using the facility drops to P_2, then its volume of use will increase to $V_{1,s}$ as a result of the limited behavioral adjustments households can make without moving or having its members change jobs.

If the price remains at this lower level for a prolonged period, however, some households may make investments in vehicle ownership or new residential locations, while others may change the locations where they work, shop, or engage in other activities. Each of these longer-run adjustments increases use of the transportation network at this lower generalized price, so that travel volume on it will eventually increase further to V_2, the value of the long-run demand function evaluated at price P_2. Thus the travel-inducing effect of a decline in generalized price is likely to be larger over the long term than in the short run.

Further changes in the generalized price from P_2 will produce instantaneous movements along the short-run demand curve passing through that point, labeled D_2. If the price continues to decline, another short-run demand curve would be established still farther to the right, but if the generalized price were to return to its original value P_1, the volume of travel on the facility would decline only to $V_{2,s}$ in the short run. Over the longer run, volume would eventually decline toward its original level if the price remained at this higher level, but the durability of some investments made in response to the lower price may inhibit volume from returning completely to V_1.[3]

---

[3] Lee et al. (1999) use the short versus long-run distinction to distinguish between induced *traffic* (or travel) and induced *demand*. Immediate changes in traffic volumes are the result of movements along the short-run demand curve; but in the long run, the short-run demand curve can shift outward, meaning that a higher volume of traffic will prevail at every value of generalized price. Thus induced traffic represents 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.

Induced Traffic versus Induced Demand

The specific sources of induced travel depend on the spatial and temporal extent of the "market" being considered, as well as on this distinction between the short and long run. If travel on a specific facility or within a single corridor is represented on the horizontal axis, diversion of travel from other routes, destinations, or travel modes in response to reductions in the generalized price of using it will represent induced usage. If the use of that facility or corridor during a limited time period such as the morning peak hour is the focus of analysis, trips that are rescheduled from other hours to that time period as a result of a reduction in the generalized price of traveling then will appear to represent induced travel.

These sources of induced travel may be large relative to its level of usage before the decline in price that causes these diversions. Nevertheless, they are short-run adjustments to the lower generalized price because they can occur without changes in residential and job locations or in vehicle ownership. Thus in the context of Figure 3, they would be represented as movements along the short-run demand curve associated with a fixed point along the long-run demand function, which Lee *et al.* (1999) refer to as induced *traffic* occurring in response to reductions in the generalized price of travel on specific facilities or during limited times.

Where the question is whether expanding transportation system capacity generates entirely new travel within the region it serves, however, its likely sources are new, more frequent, or longer trips. These would result from increased participation by household members in activities outside the home, changes in their residential locations or in their members' employment locations, or adjustments in the number and types of vehicles they own, all of which are inherently longer-run responses. Any of these responses would be reflected in movements along the long-run demand function for total travel on the transportation network or system serving the region, and the establishment of a new short-run demand curve exhibiting higher travel volumes at each generalized price. Lee *et al.* reserve the term induced *demand* for these responses.

Short and Long-Run Responses to Expanding Capacity

Figure 4 combines Figures 2 and 3 to illustrate the distinction between the short and long-run responses to expanding the capacity of a transportation system. As in Figure 2, an investment that increases the capacity of a transportation network or system shifts the price function relating generalized cost per user to its level of usage downward or outward, or from its initial position PF_1 to PF_2. As Figure 3 indicated previously, this initially causes a downward movement along the short-run demand function D_1, so that the decline in generalized price per user from its initial value P_1 to $P_{1,s}$ leads to an expansion of travel from V_1 to $V_{1,s}$. Over time, however, as the various behavioral adjustments available to households in the longer term increasingly occur, usage of the system expands further to V_2, where the price function associated with its expanded capacity intersects the long-run demand function for the transportation services it provides.

In Figure 4, it is important to note that induced demand for travel on an expanded system increases the level of congestion and results in a higher post-expansion generalized price than would occur if demand did not respond. Thus even in the short run, the upward slope of the price function – a consequence of the effect of congestion on travel speeds – results in an equilibrium price $P_{1,s}$ that is higher than that if there were no short-run response of demand to declining generalized prices (that is, if the short-run demand curve D_1 were vertical). In the longer term, the upward slope of the price function also results in a final equilibrium price P_2 that is above its interim value $P_{1,s}$, although still below its pre-expansion level P_1. Even in the long run, however, congestion on the expanded network and the resulting generalized price P_2 *cannot* return to the same level as prevailed before the expansion as a result of induced demand alone.[4]

---

[4] Economists are still searching for the first instance of a long-run demand curve that does not slope downward. In any case, if the demand curve for travel on a regional network were perfectly flat – that is, if demand were infinitely sensitive to price – any increase in the generalized price of travel would eliminate all travel, which seems extremely unlikely.

The Significance of Induced Demand

Most controversy arises over the extent of this entirely new travel resulting from reductions in the generalized price caused by investments in expanding transportation infrastructure. Empirical estimates of the relative importance of the short-run ($V_{1,s}$ minus V_1 in Figure 4) and longer-term (V_2 minus $V_{1,s}$) effects of lower generalized prices for travel vary widely. One reason for this is that disentangling the long-run effect of improved travel speeds and lower generalized prices on transportation system usage from exogenous changes in the demographic and macroeconomic factors that also affect travel demand is extremely difficult. Compounding this difficulty is the fact that opportunities to measure the response of demand to generalized price changes arise primarily where investments in expanding network capacity are made, which tend to be exactly where demographic and macroeconomic growth is most rapid and thus most likely to confound measurement of the response to price.

In terms of Figures 3 and 4, the long-run demand curve D_{LR} tends to be moving outward as a result of regional demographic and economic growth most rapidly in locations where investments in new or expanded transportation system capacity -- made in response to exactly those same forces -- are also "pushing" the short-run demand curve D_1 down along the long-run demand function. The result is that it is extremely difficult to isolate the effect on travel demand from the reduction in generalized prices that occurs in response to an investment in expanded capacity from increases in demand due to the demographic and regional economic growth that led public officials to expand transportation system capacity. This difficulty – which presents a conceptual dilemma as well as a measurement problem – turns out to be the source of much of the controversy over the significance of induced demand.

A Few Details

Finally, two miscellaneous points about induced demand are worth noting. First, induced demand is not unique to highways; demand for all forms of transportation services increases in response to reductions in their generalized prices. Thus for example investments in new rail transit lines can lead to exactly the same changes in residential and employment locations -- and resulting increases in travel demand -- as do highways, at least where they improve travel speeds in a corridor significantly. Second, the nature of demand for the services provided by transportation infrastructure that is used primarily by operators of commercial transportation services – airports, port facilities, railroads, inland waterways – is somewhat different than household demands for personal travel, since these services are inputs into commercial transportation operators' production processes.[5] Nevertheless, they increase in response to declines in the generalized price of these services, exactly as households' demands for the services provided by highways do.

Measuring Induced Demand

The quantity we want to measure is any increase in total travel over an entire network or transportation system that occurs in response to an investment in expanding its capacity. More specifically, we want to focus on the increase in network or system usage resulting from entirely new travel – new trips or travel to more distant destinations – that is spurred ("induced") by households' and firms' longer-run behavioral responses to the reduction in the generalized price function for travel that results from the investment in expanding the system's capacity. These responses can include changes in households' residential locations or in their members' workplace locations, increased participation by household members in activities outside the home (and thus requiring travel to reach), rescheduling of current activities to more desirable times of the day, and increases in the number of vehicles they own or use.

---

[5] See Walter Nicholson, *Microeconomic Theory: Basic Principles and Extensions*, 7th edition, Dryden Press, 1998, Chapter 23, for the author's usual very readable discussion of firms' demands for factors of production. In this context, induced demand has a convenient interpretation: it is simply the output effect of a reduction in the price services provided by transportation infrastruc-

What Are We Trying to Measure?

Referring back to Figure 4, the quantity of central interest is equal to V_2 minus $V_{1,s}$, the long-term increase in travel that occurs in response to a shift in the generalized price function beyond that resulting from movement along the short-run demand function. The causal chain that ultimately produces this response is important to bear in mind: it is set in motion by an investment that alters the physical design attributes of some part of the network, such as its width, curvature, grade, or surface condition. In turn, these changes interact with driver behavior to cause travel speeds on the improved part of the network to decline more slowly as its usage increases, and thus to be higher at every level of usage than previously (at least up to the legal speed limit).[6]

The degree of improvement in travel speeds and other conditions resulting from an investment that expands its capacity depends on many factors, including the design of the facility, the specific features that are changed by the investment, and its pre-improvement level of usage. Whatever its degree, however, since travel time is an important element of the generalized price faced by travelers using the network, the result is the outward or downward shift in the generalized price function from PF_1 to PF_2 shown in Figure 4. As the figure illustrates, this shift in the generalized price function produces the short-run increase in travel from V_1 to $V_{1,s}$, as well as the further increase to V_2 that follows over the longer term.

The Elasticity of Travel Demand

Economists commonly measure the sensitivity of demand for a good or service to changes in its price using a dimensionless parameter called the price elasticity of demand. There are several different versions of the elasticity measure, but the most common (called the "point elasticity") is the percent change in

---

ture (such as highways) on demand for those services by firms and households who employ them as an input to the process of producing transportation services.

the quantity of a good or service that is demanded (purchased) in response to a one percent change in its price. Because rising prices reduce the quantity demanded, while declining prices stimulate an increase in quantity demanded, price elasticities have negative values.

The magnitude of induced demand for travel on a network or system that occurs in response to an investment in expanding its capacity depends on both the resulting decline in the generalized price of travel and the elasticity of travel demand with respect to this generalized price. Referring again to Figure 4, the decline in price from P_1 to P_2 produces an increase in travel from V_1 to V_2, of which we are most interested in the component from $V_{1,s}$ to V_2, and these changes can be expressed in percent or proportional terms by dividing them by the initial price and volume of travel. Although it is difficult to estimate the final equilibrium value of P_2 because of the movement along the new price function as the volume of travel increases from V_1 to $V_{1,s}$ and ultimately to V_2, it may be possible to simulate or otherwise estimate the percentage by which the price function shifts downward at the original volume of travel V_1.

Combining this with estimates of the short-run (generalized) price elasticity of demand would yield an estimate of the percent increase in travel occurring in the short term (previously referred to as induced *traffic*), or the movement from V1 to $V_{1,s}$. A similar calculation would yield an estimate of the long-run percent or proportional increase from V_1 to V_2, and we could calculate the value of V2 minus $V_{1,s}$ – previously defined as the induced *demand* resulting from the downward shift in the price function -- as the difference between this and the previous quantity. One problem with this approach, however, is that we may not have empirical estimates of the elasticity of demand with respect to generalized price available, for either the short or long run.

While the generalized price is a useful abstract concept, what we actually observe and may be able to measure are movements in its individual compo-

[6] They my also cause other elements of generalized price to rise less rapidly with increasing us-

nents, such as vehicle operating costs, road tolls, or travel time, and we may still be uncertain about the monetary value of the last of these. We can solve this problem by using the fractions of generalized price that each of these components represents to convert these to equivalent movements in the generalized price. Another possibility is to estimate separately the response of travel demand to each component of generalized price that is affected by a capacity expansion using a demand elasticity measured with respect to that component alone, and then to add our estimates of these individual components of induced demand using the component shares as weights. Lee (1999) provides empirical estimates of component shares of the generalized price for automobile travel, as well as of demand elasticities for each of these components.

The single best index of the likely magnitude of induced demand resulting from investments in expanded network capacity is probably the estimated value of the long-run elasticity of travel demand with respect to travel time per unit of distance (the reciprocal of speed). What is important to avoid is the understandable temptation to subsume the response of travel demand to investments in capacity expansion in a single parameter that attempts to measure the combined effect of investment on the generalized price of travel and the response of demand to changes in its generalized price. Referring again to Figure 4, such a parameter would have to measure the magnitudes of the downward shift in the price function resulting from an investment, the sensitivity of travel demand to the resulting decline in the generalized price of travel, and the "steepness" of the new price function over the range from V_1 to V_2. The resulting opportunities for measurement error and difficulty in interpreting the value of such a measure make this approach fraught with difficulty, yet this is exactly what the most commonly-used measure of induced demand attempts to do.

---

age of the network.

The "Capacity Elasticity" of Travel Demand

In empirical research on induced demand, the most frequently used measure of its magnitude is the capacity elasticity of travel demand. This parameter is defined as the percent increase in some measure of travel demand – most commonly total VMT on within a geographic unit such as a county or state – associated with a one percent increase in the capacity of its highway network, usually measured by total lane-miles. It measures the response of usage of a highway network (or some part of it, such as all facilities of a certain functional class such as freeways) to changes in its capacity resulting from investments that extend the network or widen its existing links.

Because higher-capacity facilities enable faster travel, which in turn reduces the travel time component of the generalized price of travel, this elasticity is expected to have a positive value. Some empirical research (see Noland, Fulton et al., Noland and Cowart, and Strathmann et al., for example) attempts to estimate separate short and long-run values of the capacity elasticity of demand. As the previous discussion indicated, the long-run estimate of this parameter is expected to exceed its short-run value. The short-run elasticity is a measure of induced *traffic* as defined by Lee, while the difference between its estimated long and short-run values corresponds to his definition of induced *demand*.

The capacity elasticity of travel demand can be decomposed into the product of two other parameters: (1) the elasticity of demand for travel with respect to the time required to travel a unit of distance (or to its reciprocal, travel speed); and (2) the elasticity of travel time per unit distance on a highway facility or network with respect to its capacity. (See the Appendix for this derivation.) The first parameter measures the increased use of highway services that occurs in response to the higher travel speeds that result from an expansion of highway capacity.[7] As indicated previously, this is one component – probably the largest

[7] Increases in capacity may also reduce the value of vehicle services or of some vehicle operating inputs required to "produce" highway trips. The response of demand for highway services to these changes is exactly analogous to the response to increased travel speeds analyzed here.

one -- of the elasticity of travel demand with respect to its generalized price, and the difference between its long and short-run values is probably the best single index of the significance of induced demand.

The second component of the capacity elasticity, the elasticity of travel time or speed with respect to the capacity of a highway facility or system, measures the effect of investments that alter a highway's design characteristics (such as number of travel lanes, lane and shoulder width, curvature, grade, or surface condition) in ways that expand its vehicle-carrying capacity on travel time or speed. While there may be some difficulty in estimating the effect of a specific investment on travel speeds, by far the most important problem in measuring this elasticity is that its value is extremely sensitive to the level of congestion on a facility or network before such an investment is made.

At one extreme, speeds on a heavily congested facility may be improved dramatically by increasing its capacity, in which case the elasticity of speed with respect to the increase in capacity will be large, and the capacity elasticity of demand will also be large as a result. On the other hand, adding capacity to a minimally congested facility may not improve travel speeds much at all, since at most times they are already constrained by speed limits or driver behavior. Thus the resulting elasticity of travel speed with respect to capacity will be small, as will the estimated elasticity of demand with respect to capacity; in the extreme case where adding capacity leaves speeds unaffected, both will be zero.

Table 1 illustrates the wide variation in the elasticity of travel speed on a highway facility with respect to changes in its capacity. It shows the change in travel speed that results from various percentage expansions of capacity when investments in expanded capacity are made at different initial levels of congestion on the facility, as measured by different initial values of its volume-to-capacity ratio. As the table shows, the elasticity of travel speed with respect to capacity can vary widely depending on both the initial level of congestion on the facility or network where capacity is expanded (as measured by the ratio of volume to capacity before the expansion), and the proportion by which an investment actually expands its capacity.

As a consequence, the value of the capacity elasticity of demand that would be estimated from an investment in expanding the capacity of this sample facility would show a similar degree of variation, *even when the value of the elasticity of demand with respect to travel speed or generalized price is held constant*. Table 2 illustrates how the capacity elasticity of demand varies in response to the initial level of congestion on a facility or system that is expanded (again measured by the pre-expansion volume to capacity ratio) for the range of travel time elasticities reported by Cohen (2001) in his companion paper for this Forum. As it illustrates, the widely-reported capacity of demand is likely to be an extremely unreliable measure of the magnitude of induced demand, since it is so sensitive to the travel conditions initially prevailing on the facility or system where an investment in additional capacity is made.

What **Does** *the Capacity Elasticity Measure?*

In effect, the elasticity of highway travel speed or generalized price with respect to capacity is a measure of how effectively the highway planning process directs investments in expanded capacity to those facilities or parts of a network where they will produce the largest improvements in travel speeds. Empirical estimates of the capacity elasticity of travel demand will *unavoidably* subsume both the elasticity of travel speed with respect to capacity and the elasticity of demand for highway travel with respect to travel speed. Econometric estimates of the capacity elasticity will invariable suffer from an unavoidable identification problem, since the extent to which they measure the response of travel demand to increased speeds cannot be disentangled from the extent to which they capture the response of capacity investments to high demand for particular facilities or parts of a highway network.

Most of the variation in estimates of the capacity elasticity of demand will be caused by differences in the effect of capacity increases on travel speeds among different facilities or parts of a transportation network, rather than by rather than by variation in the underlying elasticity of travel demand with respect to travel time or generalized price, or by difficulties in measuring the latter parameter. In particular, high values for the capacity elasticity of demand are likely to reflect the fact that investments in new or expanded highway capacity were made in parts of the network where they improved travel speeds significantly. Although the presence of congestion by itself is not necessarily a reliable index of the value of expanding capacity to improve travel speeds, expanding the system where it is already heavily utilized has been the traditional focus of the transportation planning process. Thus high values of the capacity elasticity of demand are likely to indicate that this process is working as intended, and should certainly *not* be interpreted as evidence that the investments producing increased travel demand were undesirable.

Induced Demand and the Evaluation of Investments in Transportation Infrastructure

Much has been said and written about the implications of induced demand for the desirability of investments in transportation infrastructure, particularly highways. Even before Britain's Standing Advisory Committee on Trunk Road Assessment (SACTRA, 1994) issued its celebrated assessment of the influence of road investment on traffic levels, its Royal Commission on Environmental Pollution (1994) had recommended that the nationwide level of capital investment in new and expanded highways planned for the next two decades be cut by half, with only the simple assertion that 'new roads generate traffic' as its justification. And here in the U.S., the environmental community now routinely questions the desirability of investments in all forms of transportation infrastructure, with the conspicuous exceptions of urban rail transit and intercity rail lines. Their basis for doing so is that the resulting induced travel is certain to degrade environmental quality, and likely even to make these investments self-defeating from a transportation standpoint because congestion will soon return to its original level.

Treating Induced Demand Properly

This is an extremely curious situation, which I believe arises more from a misplaced focus on the demand-inducing effect of investment than from a failure to treat it properly in evaluating proposed infrastructure investments, although both of these factors undoubtedly play a role. The main reason I find this argument so troubling is that recognizing the sensitivity of demand for use of a transportation system to changes in travel speeds or other elements of the generalized price of travel introduces an additional category of benefits from investments in expanding its capacity. This occurs because any additional travel that is induced by the improvement in travel conditions resulting from the investment provides significant benefits to the households or businesses whose more frequent or longer trips account for it. The magnitude of these benefits depends directly on the magnitude of induced demand resulting from the improvement in travel speeds or reduction in costs, and the greater is induced demand, the *larger* are the additional benefits from recognizing it.

At the same time, however, induced traffic increases the level of congestion that would otherwise occur on an improved network and thus offsets some of the travel time savings to those using it before it was improved. Whether this reduction in benefits to previous users of the network is large enough to offset the additional benefits from new travel is an extremely complex question, the answer to which depends on three factors:

- the elasticity of total travel demand over a facility network with respect to the generalized cost of traveling on it;
- the relationship of travel speed and other elements of the generalized price of travel to the level of usage of the facility or network after it is improved;
- the effect that the proposed investment would have on travel speed and other elements of generalized cost if demand did not increase in response to these impacts.

The specific answer is extremely sensitive to the exact value of demand elasticity and the sensitivity of generalized price to usage (which is similar to an elasticity of supply for the facility or network, but not quite the same), and it is difficult to identify a general case or rule.

Figure 5 illustrates the effect of recognizing induced demand on the benefits from an investment that expands the capacity of a transportation facility or network. As in Figure 2 previously, the investment shifts the price function for the facility from PF_1 downward to PF_2, so that the generalized price of travel is lower at any volume of travel. If demand did not respond to this reduction in price, volume would remain at its original level of V_1, while the generalized price (per mile, trip, or other measure of usage) of travel would decline to P_0. This would result in a price reduction of (P_1-P_0) for each user or trip on the improved system, and thus in total benefits of V_1 times (P_1-P_0).

However, the response of demand to the reduction in generalized price results in an increase in usage of the improved system to V_2, which raises the generalized price of traveling on it from P_0 to P_2 because the induced demand -- equal to V_2 minus V_1 -- produces some congestion on the improved facility. The benefits to new trips or users of the improved system are equal to the difference between what value to travelers, which is measured by the height of the demand curve, and the price actually paid for them, or P_2. Thus the additional benefits resulting from the recognition of induced demand collectively amount to the triangle-like area abc, which is usually approximated by the quantity $(1/2)(P_1-P_2)(V_2-V_1)$. At the same time, however, the increase in travel from V_1 to V_2 -- the induced demand for use of the improved facility -- reduces the benefits received by each pre-improvement user by the amount (P_2-P_0), and in total by the area V_1 times (P_2-P_0).

The net effect of incorporating induced demand into the evaluation of benefits from the investment in additional capacity is thus equal to the difference between the values of $(1/2)(P_1-P_2)(V_2-V_1)$ and $V_1(P_2-P_0)$. As the elasticity of travel demand with respect to its generalized price increases in magnitude, the quantity

(V_2-V_1) grows, which by itself would cause this quantity to increase. However, so does the value (P_2-P_0), because of the relationship of travel speed to usage of the improved facility represented by the price function PF_2, and this has the effect of reducing the benefits from induced demand. Finally, the value of (P_1-P_0) is determined by the size of the investment and its consequent effect on travel speed if demand did not increase in response, but the relative size of the components (P_1-P_2) and (P_2-P_0) into which it is partitioned depends on *both* the elasticity of demand and the behavior of the price function.

Thus all three factors combine to determine the overall effect of recognizing induced demand on the benefits from an investment that expands capacity. Unfortunately, however, the specific relationship between these benefits and these three parameters can be complex even for relatively simple forms of the demand and price functions. Where the demand and price functions are linear (or where the changes in price and travel volume prompted by the investment are small enough to make this a reasonable approximation), for example, show that the change in benefits from recognizing induced demand is equal to

$$\Delta B = [\varepsilon_D(1-P_0/P_1)]/[2(1+\varepsilon_D\varepsilon_P)^2] - [(\varepsilon_D\varepsilon_P)/(1+\varepsilon_D\varepsilon_P)]$$

where ε_D is the elasticity of travel demand with respect to its generalized price, ε_P is the elasticity of the generalized price for travel on the improved facility with respect to the volume of usage it experiences, and P_1 and P_0 are – as in Figure 5 -- the generalized price originally prevailing on the unimproved facility and the generalized price that would occur on the expanded facility if demand did not increase in response to the decline in price. Thus the quantity $(1-P_0/P_1)$ – or written another way, $(P_1-P_0)/P_1$ -- is a measure of the proportional reduction in which generalized price that would occur as a result of the capacity expansion if there were no response of demand.

Williams and Moore (1990) provide detailed estimates of the effect on benefits from recognizing induce demand – or viewed another way, the error in estimating benefits when induced demand is ignored – under different assumptions about the form of demand and price functions and a wide range of elasticity values. Their results confirm that the effect of considering induced demand on the benefits from capacity expansion is very sensitive to the elasticity of demand, initial congestion level, behavior of the price function for the improved facility, and size of the capacity expansion.[8] With demand elasticities in the range of those reported by Cohen (-0. 2 to –0.4), incorporating induced demand will increase benefits slightly if initial congestion levels or the elasticity of generalized price with respect to usage of the new facility is low. Recognizing induced demand will reduce benefits if congestion levels are initially high, but will only do so significantly (i.e., by more than about 10%) if the facility is extremely congested both before *and* after it is expanded, so that usage of the expanded facility occurs in the range where its price function (PF_2 in Figure 5) is very steeply sloped.

One definitive conclusion is that it is *not* possible for congestion on an expanded facility to return to its original level (and thus to eliminate all benefits from the investment) as a consequence of induced demand alone. The reason this situation seems to occur is probably that investments in expanded capacity are most commonly made where demand for travel on a facility or in a corridor is growing most rapidly in response to demographic or economic growth, so that congestion levels would have increased even more rapidly in the absence of investments in expanded transportation system capacity. The accompanying argument that because expanding capacity will simply cause congestion to return to its original level, it produces no benefits, also reflects a failure to specify the proper "counterfactual" case in project evaluation. A correct evaluation compares future travel conditions with and without the proposed investment, not travel conditions after the investment to those before it is made, and explicitly

[8] One exception is where an investment eliminates congestion on a facility, in which case recognizing induced travel increases total benefits regardless of the elasticity of demand (because there is no erosion of time savings to previous users), but it is unclear how common this situation is likely to be.

recognizing induced demand within this framework will generally yield additional benefits.

Externalities and Induced Demand

In my view, the recent controversy over the effect of induced demand on the desirability of investments in expanding transportation facilities arises from the unpriced externalities caused by transportation, particularly private motor vehicle travel, rather than from the existence of induced demand itself. I believe that the potential *increase* in these external costs – which include the delays travelers impose on one another, health and property damages caused by vehicles' contributions to air pollution, the effects of vehicle noise, and part of the costs of transportation accidents – associated with induced demand is the real reason to be concerned about it when evaluating proposed investments in transportation infrastructure. At the same time, it is also important to keep in mind that most investments in increased transportation system capacity are likely to reduce the magnitude of at least some of these externalities, particularly congestion delays, air pollution damages, and accident costs.

The underlying problem is that travelers are not charged for the for the value of congestion delays, air pollution and noise damages, or accident costs they impose on other transportation system users and on the public at large. In deciding how much to travel, travelers weigh the benefits from making each trip against only those costs they incur – the value of their own travel time, costs to operate their own vehicles, and so on – and quite reasonably ignore the costs they impose on others. Thus before the capacity of a transportation network is expanded, the value of some trips will be less than the full costs they impose on society as a whole (including travelers and the general public), even though their value exceeds the costs borne privately by the travelers who make them. The collective excess of costs imposed by these trips over the benefits they provide to travelers is properly treated as a net cost that should be deducted from the net benefits provided by the existing or "baseline" transportation system.

Figure 6 illustrates the situation where transportation activity generates significant externalities, such as congestion, emissions of air pollutants, or noise. While the price function shows how the generalized prices faced by individual travelers varies as usage of the transportation system increases, it does not capture the costs that these externalities impose on other users or on the public at large. Each user's contribution to increased congestion causes delays to other travelers using the system during the same time period, while motor vehicles' contributions to air pollution and noise cause health and property damages that can impose significant costs on the general public.

The price function for each level of capacity is associated with a marginal cost function that includes not only the components of travel cost that are borne by each user, but also the value of these external costs that each additional (or "marginal") user imposes on other travelers and on the public. This marginal cost function lies above the price function associated with that same level of system capacity, reflecting the fact that the total costs imposed by each additional trip or traveler are only partly borne by users themselves.[9] In Figure 6, the marginal cost function MC_1 is associated with the price function PF_1, while MC_2 is associated with the price function for the expanded facility, PF_2.

Each marginal cost function diverges progressively further from its associated price function as the level of usage increases, primarily because increased congestion causes the delays each additional user imposes on others to escalate with the level of usage. Since the effect of increased capacity is to reduce the level of congestion that results at any volume of travel, investments in expanded capacity tend to reduce the divergence between the associated price

[9] It is not necessary that the price function lie below the marginal cost function. Prices for using the transportation system, including explicit user charges such as tolls or implicit charges such as fuel taxes, could be set so that they equaled or even exceeded the sum of costs for providing infrastructure and the external costs each additional user imposes. Since the remaining components of generalized price -- such as vehicle capital and operating costs, the value of traveling time, etc. -- are necessarily borne by users, this would cause the price function to be identical to or above the marginal cost function. By far the more common situation, however, is that user charges or fuel taxes recoup at most transportation infrastructure costs, so that the price function lies below the marginal cost function.

and marginal cost functions. Thus in Figure 6, the vertical distance between MC_2 and PF_2 is smaller than that between MC_1 and PF_1 at each level of usage, at least until it approaches the maximum capacity of the expanded system.

The level of facility or system use is established by the intersection of the demand curve with the price function, rather than with the marginal cost function. Thus in Figure 6, the volume of travel on the facility before its capacity is expanded, V_1, is determined by the intersection of the demand function with PF_1, while the post-expansion volume V_2 represents the intersection of the demand function with the price function for the expanded system, PF_2. Since the height of the demand curve represents travelers' valuation of the benefits they receive from using the system, some trips will be made that provide benefits to travelers that exceed the generalized price they bear but are less than the full economic (or "social") costs they impose.

The resulting "welfare loss" is equal to the excess of costs these trips impose over the benefits they provide to the travelers who make them. Before the system expansion, it is equal to the approximately triangular area labeled abc in Figure 6, which is bounded by the demand curve, the marginal cost function MC_1, and the line corresponding to the level of use V_1. After the system's capacity is expanded, the analogous loss is the area def, which is bounded by the demand curve, MC_2, and the line corresponding to V_2. Comparing these two areas suggests that the value of this welfare loss tends to be reduced by investments in expanded transportation system capacity, although it is admittedly difficult to see this clearly in Figure 6.

It is important to note, however, that the welfare loss on this "excess" travel is *not* a consequence of induced demand, nor is the benefit from any reduction in its magnitude that occurs as a result of expanding capacity. Instead, it results from the existence of unpriced externalities generated by usage of the transportation system, together with any failure of user charges to reflect the costs of providing transportation infrastructure. Because investments in added transportation capacity do not alter the systematic underpricing of transportation system use, this situation will persist even after such an investment is made. However, expanding transportation system capacity is likely to affect travel conditions in ways that reduce the magnitude of some of these externalities, particularly congestion delays and accidents, and possibly air pollution damages as well. By doing so, it will reduce the excess value of costs imposed by travelers making these "infra-marginal" trips on the expanded system over the benefits they receive, compared to the analogous excess of costs over benefits for the baseline or unimproved transportation system.

Thus a reduction in the welfare loss arising from the failure to price transportation-related externalities represents a potential additional benefit of expanding transportation system capacity, although its magnitude is likely to be small compared to the benefits to pre-expansion users and from induced demand. It is theoretically possible that the amount by which these external costs exceed travelers' benefits from the trips causing them could increase as a consequence of expanding transportation capacity. However, this can only occur where demand is sufficiently sensitive to a reduction in the generalized price of travel that usage of the expanded system generates higher levels of some externalities than would occur on the unimproved system, which seems extremely unlikely. It is even more difficult to imagine that any increase in this deadweight loss resulting from induced demand could outweigh the combined benefits to pre-improvement users of the system and those induced to make the additional trips.

In any case, the response to the problem of externalities from transportation system use should be to adopt measures to price or limit them, rather than to forego otherwise desirable investments in an effort to curtail them. Federal emission controls have been enormously successful in reducing motor vehicles' contributions to air pollution, and at the same time, changes in road and vehicle designs have combined to reduce accident rates, fatalities, and injuries significantly. Although pricing to limit congestion – which accounts for the largest share of transportation externalities -- has not been widely adopted, congestion costs are borne collectively by travelers rather than by the public at large and are thus not "external" from the viewpoint of the transportation system.

Evaluation with Both Induced Demand and Externalities

Economic evaluation of proposed investments in new or expanded transportation facilities in the presence of both induced demand and unpriced externalities is certainly more complicated than without them, but can nevertheless be done. Referring again to Figure 6, the benefits from an investment that expands the capacity of a transportation network from the level associated with the price and marginal cost functions PF_1 and MC_1 to the capacity associated with PF_2 and MC_2 have three components: (1) the reduction in generalized price for travel that occurs before the capacity expansion, equal to (P_1-P_2) times V_1; (2) the benefits from new travel induced by the reduction in generalized price, whose value is the approximately triangular area cgf; and (3) any reduction in the welfare loss resulting from unpriced externalities, which is equal to the difference between the roughly triangular areas abc and def. Performing these calculations is clearly more difficult that simply calculating time and cost savings to pre-improvement users of the network, but various methods are available to estimate the three benefit components using observable data (see for example Lee, 2000).

The inherent "problem" is neither the difficulty of conducting sound benefit-cost evaluation of proposed investments in the presence of elastic demand and externalities, nor that we ignore induced demand when we conduct economic evaluation of investment proposals. Instead, the problem – at least as I see it – is that we evaluate proposed investments in transportation infrastructure carelessly or not at all, and too often on the basis of "benefits" such as job creation that on closer examination turn out to be costs instead. As a result, we undoubtedly make some poorly-chosen investments in our transportation infrastructure, but when we do, it is rarely because we ignore induced demand.

Induced Demand and Transportation Policy

Recognizing the sensitivity of demand to changes in travel speed or to other elements of the generalized price of travel and treating the additional demand correctly in evaluating proposed expansions of the nation's transportation infrastructure will ultimately lead to better investment decisions than will either ignoring induced demand or attempting to suppress investment as a means of controlling undesirable by-products of transportation activity. Incorporating induced demand into the evaluation of proposed infrastructure investments may raise or lower their total (net) benefits, depending on a complex set of factors. This is because induced demand not only provides an additional source of benefits (which depends on the elasticity of travel demand), but also affects the benefits experienced by previous users and the level of externalities generated by travel on the expanded system (which depend on the relationship of travel speed, accidents, and vehicle emissions to usage of the expanded system).

The complexity of these factors is a compelling reason for more careful evaluation of proposed investments, including more detailed analysis of the likely magnitudes of each of these impacts and more careful assessment of the relationships among their magnitudes. The potential for expansion of transportation systems to induce additional travel demand is *not* a reason to forego investments in physical facilities or new technologies that increase capacity. Total benefits from investments that are insufficient to justify their costs even when the effects of induced travel are explicitly included are a reason to forego them, but whether this is the case for specific investment proposals can only be determined by careful analysis. It is not an issue that can or should be settled by a sweeping indictment of expanding transportation capacity that is rooted in some vague conviction that travel is objectionable, no matter how deeply it is held.

Appendix: The Capacity Elasticity of Travel Demand

Households and firms – some of which may be transportation operators, such as trucking companies – use the services provided by highways as an input to their production of transportation services. Households produce transportation services in order to allow their members to participate in activities outside the home, and may also purchase transportation services (from transit operators or taxi companies, for example) for this same purpose. Businesses produce or purchase transportation services in order to bring raw material inputs to locations where they can be used in production processes, to move finished goods to markets where they are sold, and to transport their employees for various purposes.

Except in the rare cases where tolls are charged, highway users pay no explicit price for their use of highway services; instead, they generally pay taxes on their use of fuel, and face costs for using highway services in the form of time their members spend as auto drivers and passengers. This "time price" -- the amount of time entailed in making a trip between a given origin and destination -- is determined by the speed of highway travel. The rate at which individuals value traveling time reflects the opportunity cost of time itself, which may differ depending on whether they are engaged in personal or "on the job" travel, as well as on the disutility of spending time in an automobile. Finally, households' and businesses' demands for highway services also depend on prices of other inputs into the process of producing transportation services, including depreciation rates for vehicles, and prices for fuel and vehicle maintenance.

The quantity of highway services demanded for household and business travel during any time period is usually measured by the volume of vehicles using a highway facility or network (V). Given the number and characteristics of households and firms served by the facility or network, total demand for use of the highway will depend on the speed of travel (S), since speed determines time per unit of distance traveled, and on prices for vehicle services (p_v) and vehicle operating inputs such as fuel (p_o):

(1) $$V = v(S,p_v,p_o)$$

As travel speeds increase, the travel time component of the cost of highway travel declines, and both businesses and households will make more trips on the facility or network. Households will engage in more travel either to satisfy their members' demands to participate in more activities outside the home, to engage in current activities at more distant locations, or to substitute for trips that formerly used other means of travel that made less intensive use of highway services. Businesses will substitute increased use of transportation for other inputs into their production and distribution activities, such as by reducing inventory levels or relocating manufacturing or distribution facilities. As a consequence, we expect that

(2) $$\partial V/\partial S > 0$$

Highway Capacity and Travel Speed

The speed of highway travel depends on various characteristics of road and highway design, including the number of travel lanes, width, curvature, grade, and surface condition. For convenience, these are usually summarized by its maximum vehicle-carrying capacity (or flow rate) per time period (C). Travel speed on a highway facility or network also depends on the volume of vehicles using it at any time; thus

(3) $$S = s(V,C)$$

Substituting this into the demand function for highway services gives:

(4) $V = f(C, p_v, p_o)$

which shows the indirect dependence of total demand for the services of a highway facility or network on its capacity.

A higher-capacity highway facility or network can accommodate larger travel volumes at any speed, usually because it includes more travel lanes or because its design characteristics make drivers willing to tolerate closer vehicle spacing (or higher vehicle densities) at most travel speeds. Thus

(5) $\partial S/\partial C > 0$

The Capacity Elasticity of Demand

The capacity elasticity of travel demand (or more precisely, of demand for highway services), ε_c, is defined as

(6) $\varepsilon_c = \partial V/\partial C\ (C/V)$

It measures the response of usage of a highway facility or network to changes in its capacity, which result from investment (or less frequently, disinvestment) in modifying design characteristics -- most commonly its width or number of lanes -- that limit its capacity to accommodate vehicle flows. Because higher-capacity facilities enable faster travel, which in turn reduces the “time price” of highway trip-making, the derivative $\partial V/\partial C$ has a positive value, and

(7) $\varepsilon_c > 0$

However, the capacity elasticity of demand can be further decomposed into

$$\text{(8)} \qquad \varepsilon_c = \partial V/\partial C\ (C/V) = \partial V/\partial S\ \partial S/\partial C\ (C/V)$$

which can be rewritten as

$$\text{(9)} \qquad \varepsilon_c = [\partial V/\partial S(S/V)]\ [\partial S/\partial C(C/S)] = \varepsilon_{v,s}\ \varepsilon_{s,c}$$

Here, $\varepsilon_{v,s}$ is the elasticity of demand for highway services with respect to travel speed, which as indicated previously should be positive in sign. The parameter $\varepsilon_{s,c}$ is the elasticity of travel speed on a highway facility or network with respect to its capacity, and as above is also expected to be positive.

The Capacity Elasticity and Induced Demand

The elasticity of demand for highway services with respect to travel speed, $\varepsilon_{v,s}$ in expression (9), is a minor variant of the familiar "travel time elasticity of demand." Because demand is expressed here as a function of travel speed rather than of its reciprocal, travel time per unit of distance, $\varepsilon_{v,s}$ is presumably positive, while travel time elasticities are negative. However, they measure the same response: the sensitivity of travel demand – or more formally, of demand for the use of highway services – to the travel time component of the cost of highway travel.

Either parameter can be used to measure the increased use of highway services that occurs in response to the higher travel speeds that result from an expansion of highway capacity.[10] This increase in the use of highway services is the "induced demand" or "induced travel" commonly attributed to investments in expanded road and highway capacity. In order to estimate its magnitude, of course, we also need to estimate the increase in travel speeds that results from the addition to facility or network capacity.

---

[10] Increases in capacity may also reduce the value of vehicle services or of some vehicle operating inputs required to "produce" highway trips. The response of demand for highway services to these changes is exactly analogous to the response to increased travel speeds analyzed here.

This improvement in travel speed – and thus the size of the demand increase it induces -- depends partly on the magnitude of the parameter $\varepsilon_{s,c}$, the elasticity of travel speed on a highway facility or network with respect to its capacity. Of course, it also depends on the size of the capacity increase itself, although there may be some question about exactly how the capacity increases from different types of investments can most accurately be measured. The major problem is that the value of this elasticity is extremely sensitive to the level and the temporal pattern of congestion that prevails on a highway facility or network before an investment in expanding its capacity is made.

At one extreme, speeds on a heavily congested facility may be improved dramatically by increasing its capacity, in which case the value of $\varepsilon_{s,c}$ will be large – values well above 1.0 are easy to imagine for initially congested facilities -- and expression (9) above shows that the capacity elasticity of demand will also be large as a result. On the other hand, adding capacity to an only minimally congested facility may not improve travel speeds much at all, since at most times they may already be constrained by speed limits or driver concerns about safety. In this case the values of both the capacity elasticity of travel speed and of travel demand will be small; in the extreme case where adding capacity leaves speeds unaffected, both will be zero.

References

Cohen, Harry, "The Induced Demand Effect: Evidence from National Data," prepared for Eno Transportation Foundation Policy Forum: Working Together to Address Induced Demand, Washington, D.C., February 22-23, 2001.

Downs, Anthony, "The Law of Peak-Hour Expressway Congestion," *Traffic Quarterly*, Volume 16 (1962), pp. 393-409.

Fulton, Lewis M., Robert B. Noland, Daniel J. Meszler, and John V. Thomas, "A Statistical Analysis of Induced Travel Effects in the U.S. Mid-Atlantic Region," *Journal of Transportation and Statistics*, Volume 3 (2000), pp. 1-14.

Lee, Douglass B., Lisa A. Klein, and Gregorio Camus, "Induced Traffic and Induced Demand," *Transportation Research Record* 1659 (1999), pp. 68-75.

Lee, Douglass B., Appendix C: "Demand Elasticities for Highway Travel," and Appendix D: "Basic Theory of Highway Project Evaluation," in *Highway Evaluation and Requirements System Technical Manual*, Washington, D.C., Federal Highway Administration, 2000.

Morris, Robert L., "Traffic as a Function of Supply and Demand," *Traffic Quarterly*, Volume 31 (1977), pp. 591-603.

Nicholson, Walter, *Microeconomic Theory: Basic Principles and Exten*sions, 6th edition, Fort Worth, Texas, The Dryden Press, 1995.

Noland, Robert B., "Relationships Between Highway Capacity and Induced Vehicle Travel," paper presented to 78th Annual Meeting of the Transportation research Board, Washington, D.C., 1999.

Noland, Robert B., and William A. Cowart, "Analysis of Metropolitan Highway Capacity and the Growth in Vehicle Miles of Travel," paper presented to 79th Annual Meeting of the Transportation research Board, Washington, D.C., 2000.

Standing Advisory Committee on Trunk Road Assessment (SACTRA), *Trunk Roads and the Generation of Traffic*, London, U.K., Department of Transport, 1994.

Strathmann, James G., Kenneth J. Deuker, Thomas Sanchez, Jihong Zhang, and Anne-Elizabeth Riis, "Analysis of Induced Travel in the 1995 NPTS," Center for Urban Studies, Portland State University, June 2000.

Williams, Huw C.W.L., and Laurence A.R. Moore, "The Appraisal of Highway Investments Under Fixed and Variable Demand," *Journal of Transport Economics and Policy*, Volume 24 (1990), pp. 61-82.

Figure 1

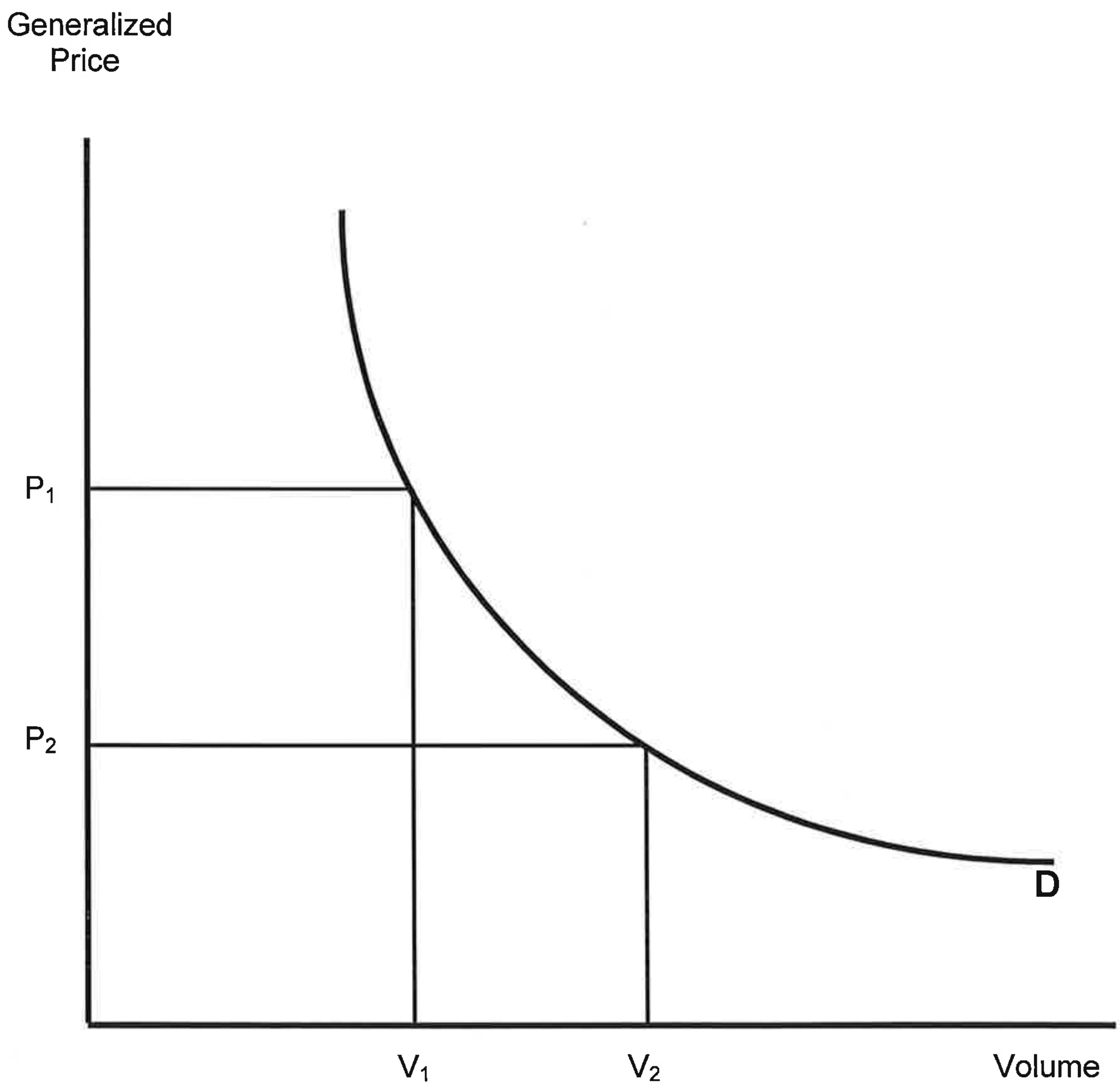
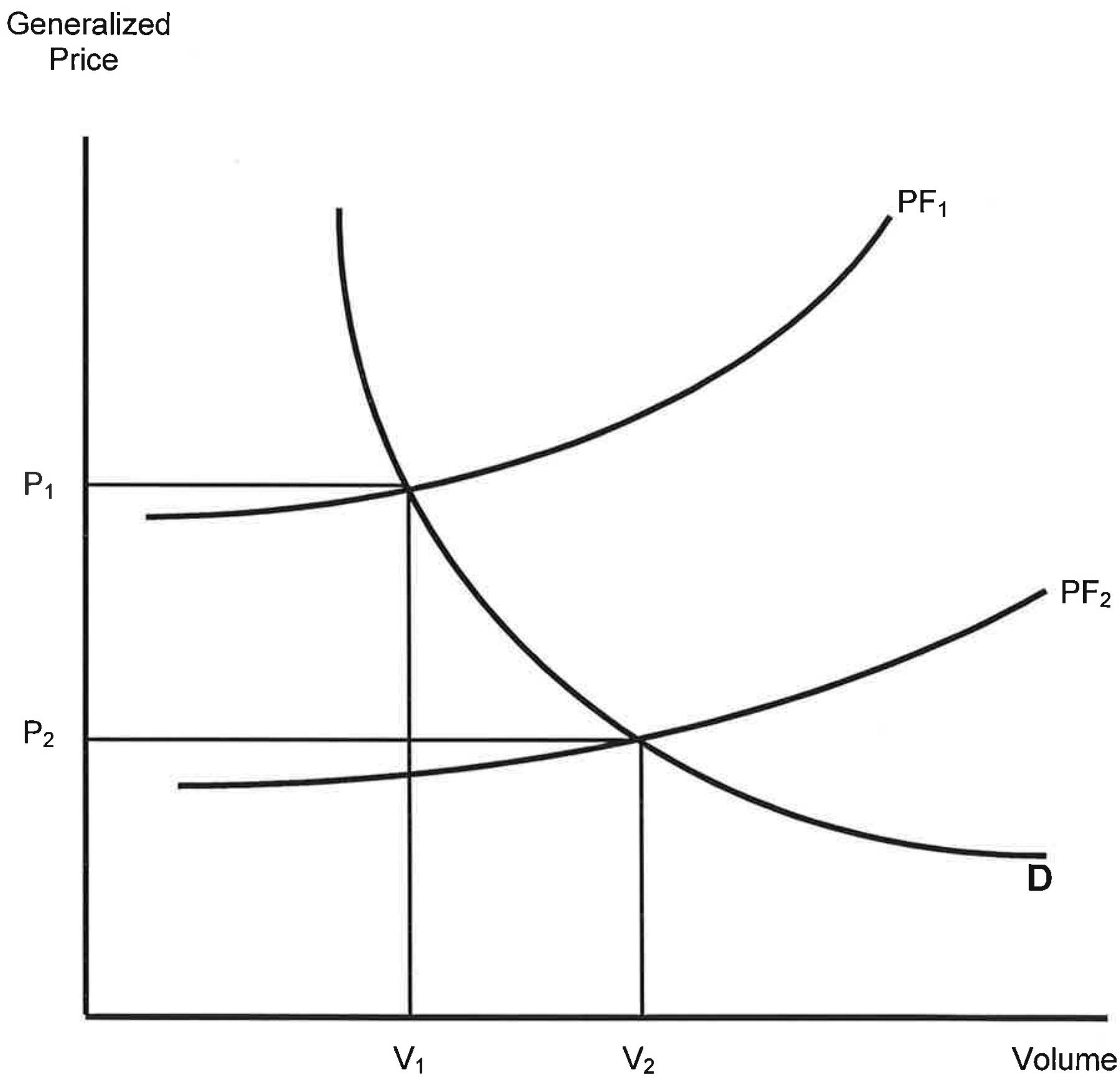


Figure 2



Don Pickrell
Draft 3/27/2001

Figure 3

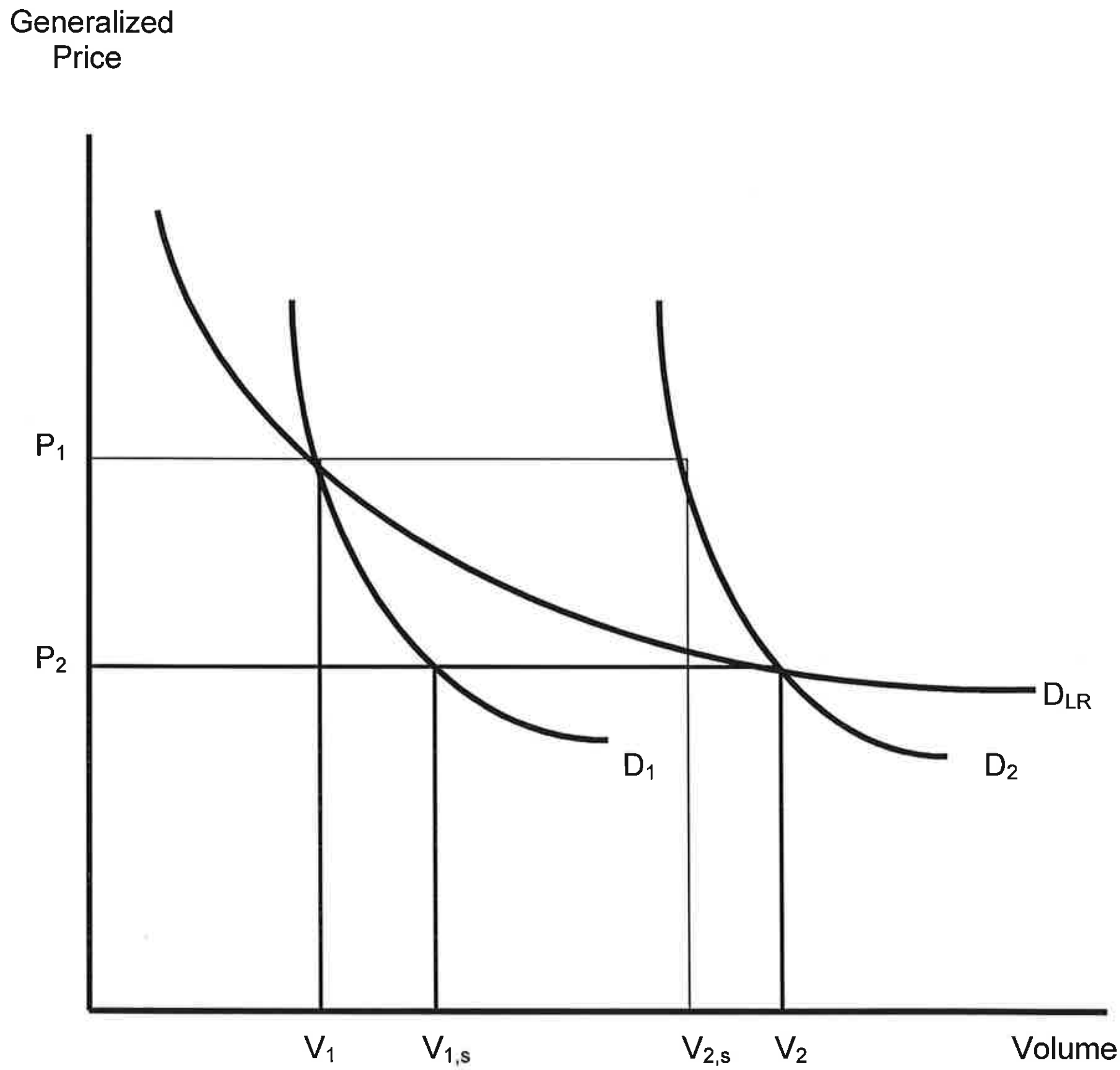


Figure 4

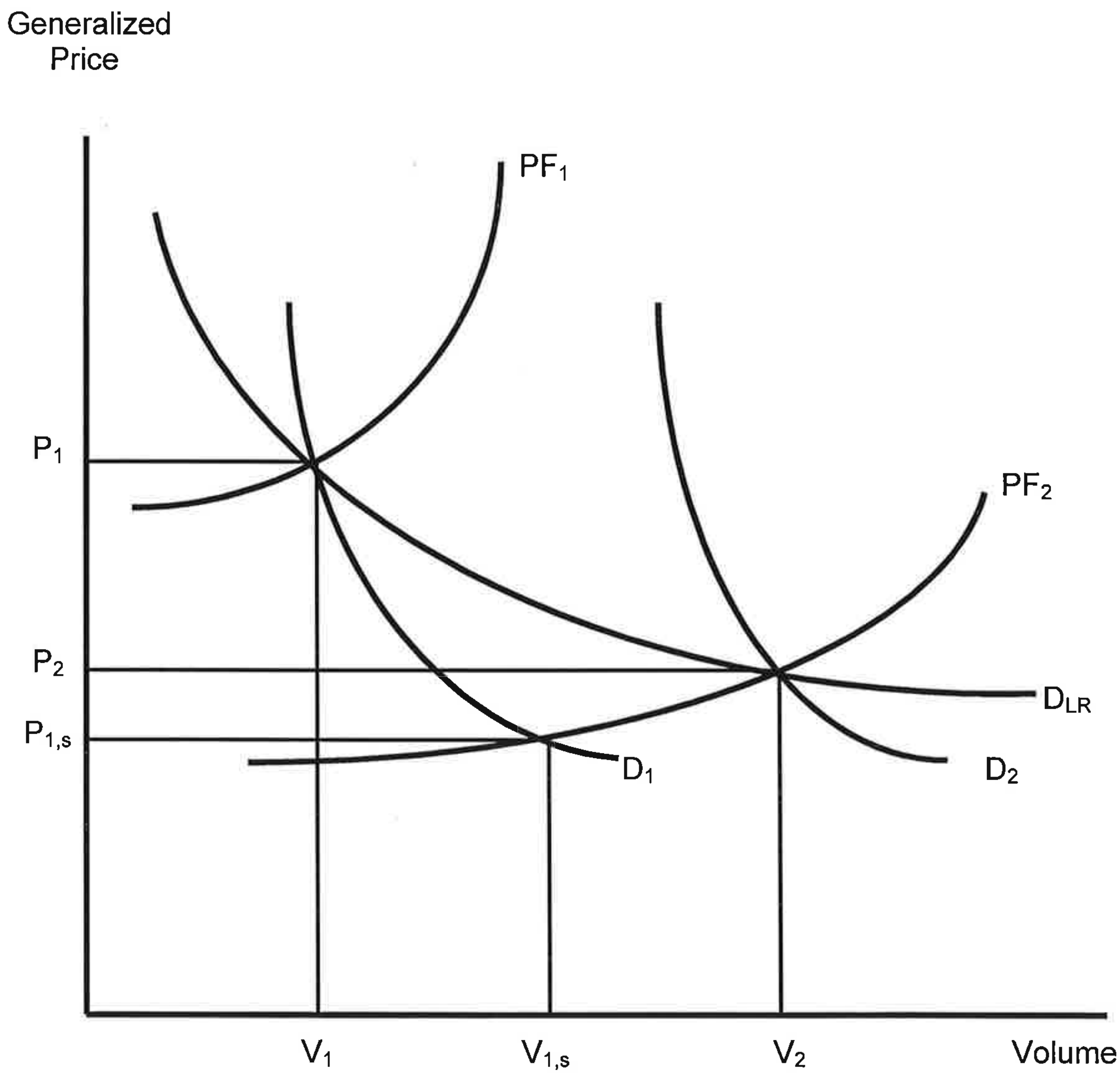


Figure 5

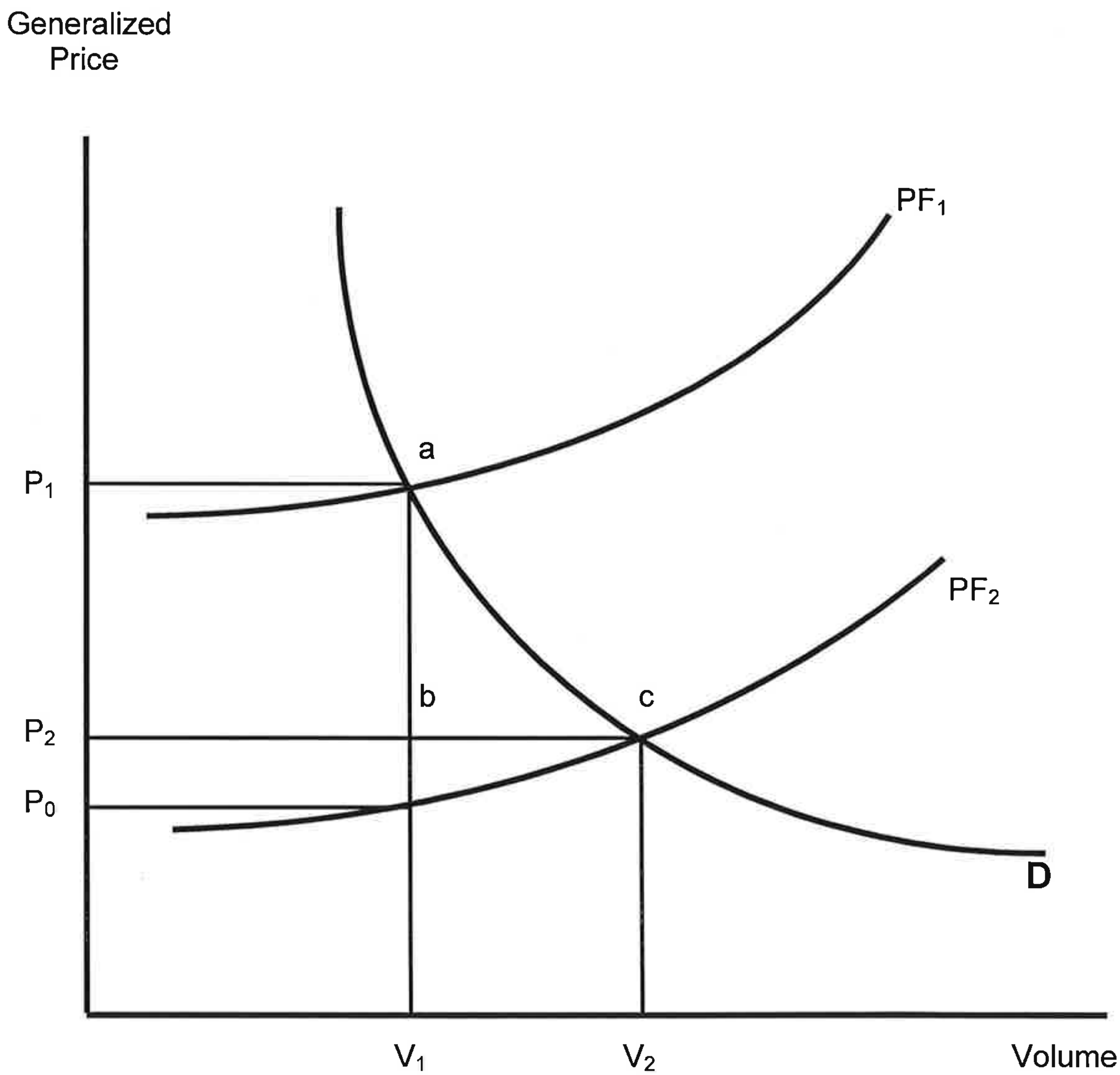


Figure 6

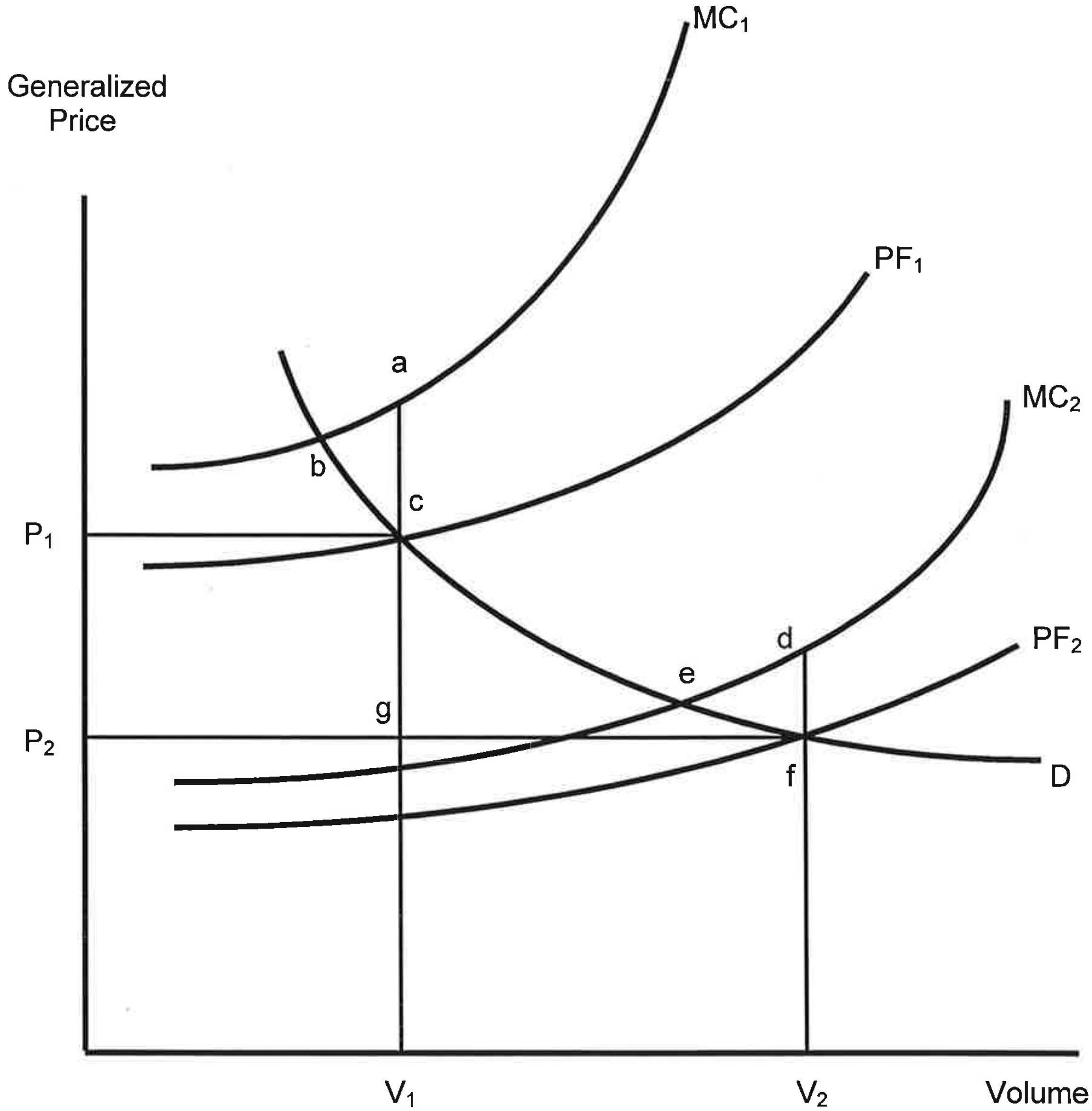

MC1
Generalized Price
PF1
a
MC2
b
c
P1
d
PF2
e
g
P2
f
D
V1
V2
Volume

Table 1
Elasticity of Speed with Respect to Capacity at Various Initial Volume/Capacity Ratios

Volume/Capacity Ratio Before Expansion	Speed (mph)	Elasticity
70%	65.0*	0.67
75%	65.0*	0.83
80%	63.0	1.02
85%	58.5	1.35
90%	53.3	1.97
91%	52.0	2.19
92%	50.8	2.47
93%	49.0	3.01
94%	47.0	4.04
95%	45.5	5.53

* Speed assumed constrained by legal limit.

Source: computed from $V/C = a_1 - a_2(S-S_0)^{a3}$, using a_1=0.95, a_2=0.001923,a_3=2.0, and S_0=55. See Small (1992) and McShane and Roess (1990).

Table 2
Relationship of Capacity Elasticity of Demand to Demand Elasticity with Respect to Speed and Initial Volume to Capacity Ratio

Volume/Capacity Ratio Before Expansion	Elasticity of Travel Demand with Respect to Speed:				
	0.20	**0.30**	**0.40**	**0.50**	**0.60**
70%	0.13	0.20	0.27	0.34	0.40
75%	0.17	0.25	0.33	0.41	0.50
80%	0.20	0.31	0.41	0.51	0.61
85%	0.27	0.40	0.54	0.67	0.81
90%	0.39	0.59	0.79	0.98	1.18
91%	0.44	0.66	0.88	1.10	1.32
92%	0.49	0.74	0.99	1.24	1.48
93%	0.60	0.90	1.20	1.51	1.81
94%	0.81	1.21	1.61	2.02	2.42
95%	1.11	1.66	2.21	2.77	3.32

Source: computed from Appendix expression (9) and elasticities of speed with respect to capacity reported in Table 1.