INTRODUCTION TO COST-BENEFIT ANALYSIS

P A R T I V

APPLICATIONS TO HIGHWAY PROJECTS

Arnold C. Harberger

University of California, Los Angeles

January 2009

Background

The key element to recognize not just with respect to highways but to many other categories of projects as well, is that one must try to focus on what is the essential benefit that the project confers. Once one takes this as one's task, it yields a pretty easy answer in most cases. Certainly this is true for highway projects, and for transportation projects generally. To say that a transportation project enables us to move people and/or goods from point A to point B has to be at best a huge exaggeration. Right now, without the project, one can get people and goods from just about every place on earth to every other place. What transportation projects to do is make it cheaper to transport people and goods. They don't enable it because it is already possible.

This simple, almost semantic step is tremendously powerful in focusing our analysis of costs and benefits, for it tells us to ask by how much a given project will reduce the costs of transport. This in turn leads to an extremely simple frame in which to place the analysis. First we have the existing traffic going, say, from A to B. We are interested in the volume of this traffic right now, not on its own merits but as a base for projecting the future path of that traffic,

if "our" project is not undertaken. Call that traffic volume V_{ot}, the amount of traffic that we expect would exist at time t, in the absence of our project. This estimate is important because it provides the base for a major component of project benefits -- namely $V_{ot}(C_{ot}-C_{pt})$. This tells us the saving in cost that we expect to be present at time t -- C_{ot} is the cost of a trip in the absence of our project, and C_{pt} is the corresponding cost in its presence. The difference $(C_{ot}-C_{pt})$ is the cost saving per vehicle, or per trip.[1]

The next major component of project benefits comes from newly-induced traffic -- that is, traffic that would not exist in the absence of the project, and hence is seen as being <u>induced</u> by the project. So if V_{pt} is the volume of traffic anticipated for time t in the presence of the project, the amount of induced traffic at that point in time will be $(V_{pt}-V_{ot})$.

The benefit that we assign to this induced traffic is $(1/2)(V_{pt}-V_{ot})(C_{ot}-C_{pt})$. Obviously, if we were to give the new traffic the same benefit as the old, we would have $(V_{pt}-V_{ot})$ $(C_{ot}-C_{pt})$. So the question is, why do we take only half of that amount? The answer, first of all, is that if someone is willing to pay more than C_{ot} to make a trip, it is taken for granted that that person will already be taking that trip in the absence of the project. Thus C_{ot} represents a

---

[1] This exposition simplifies what would ideally be done in practice on a highway improvement project. For example, one would typically have data, not only on the amount of traffic flow in the current and past years, but also on its composition. Thus, we probably would have data on heavy trucks, light trucks, buses, large, medium and small cars, etc. Probably we would make the assumption that the percentage of each vehicle type in the total would be the same in the future as in the recent past. But if we had reasons to expect it to change in a particular way, we would base our traffic projections on that expectation. Once the projects are broken down by vehicle type, we would, of course, assign to the traffic flow V_{ot} for each vehicle type its corresponding cost saving $(C_{ot}-C_{pt})$.

maximum for the “willingness to pay”, or in this case the “willingness to bear a cost” on the part of any and all of the induced traffic. A similar line of reasoning tells us that the induced traffic

shows, simply by traveling on the road in the new situation, that they are willing to bear, at the very least, the new cost C_{pt}. The standard assumption is that the actual new traffic is spread out between the minimum "willingness to pay" of C_{pt}, and the maximum (for induced traffic), C_{ot}. It is this "spreading assumption" which assigns a gross benefit of $(C_{ot}+C_{pt})/2$ to the induced traffic. Since the cost these people actually bear ("pay") in the presence of the project is C_{pt}, their net benefit is $\{[(C_{ot}-+C_{pt})/2]-C_{pt}\}$, which works out to $(C_{ot}-C_{pt})/2$, as stated above.

The obvious next question is how one determines the costs, C_{ot} and C_{pt}. One part of these consists of costs connected with the vehicles and their operation, maintenance and repair. Project evaluators can consult tapes which tell how many miles per gallon of gasoline one can expect on different types of roads -- also how much oil, tire consumption and repair and maintenance cost per mile. Finally, they can access data on how many miles a car is likely to last, if driven exclusively on each type of road (dirt, gravel, black top, concrete, etc., also level, hilly, straight, snaky, etc.) Since these factors are given in physical terms (e.g., miles per gallon, miles per tire, etc.), they can be converted to real dollar amounts by multiplying by the expected real prices (of gasoline, tires, the vehicles themselves, etc.) that are expected to prevail in period t.

Only rarely, however, do the material costs treated above account for the majority of C_{ot} and C_{pt}. The main element of the difference between these costs is almost always the time cost borne by drivers and passengers.

Consider a typical U.S. major road improvement. It might speed up traffic on the affected road, increasing it, say, from an average 25 miles per hour to, say 50 mph. Standard estimates of driving costs for passenger vehicles now range around 40¢ per mile. It is not likely

that these would change much in a case with paved roads both before and after the project. But, taking time cost at, say $10 per hour, the time cost per vehicle mile would be 40¢ before and 20¢ after the project -- a really important change!

The example above assumes one person per vehicle (the driver), and (implicitly at least, for present U.S. conditions) that the drivers and commuters are private travelers. In a real-world analysis of a road improvement we would want to value the time of truck and bus drivers at what they are actually paid (including fringe benefits but net of taxes). One would also want to make a separate estimate for the value of time for bus passengers and for the passengers in autos and other vehicles. But probably the largest category of time cost would be that of the drivers of private vehicles, most of them commuters.

This brings us to the important question of how to place a value on commuter time, for purposes of economic project evaluation. The initial reaction of most people is to assign each driver's own hourly rate (for time actually worked) as being also the relevant value of his/her commuting time. This, it turns out, is a huge overestimate -- which cannot be reconciled with the actual behavior of commuters.

One of the key contributions to the literature on the valuation of commuter time was made by Thomas Lisco, in a Ph.D. dissertation he wrote under my direction. The methodology that he used was at the time and still remains the standard approach to valuing travel time. The basic idea is very simple; where there are two relevant ways of getting from one place to another, the one that is cheaper will typically involve more travel time. Thus, in a sense, each individual commuter faces his/her own tradeoff between travel time on the one hand and money on the other. In Lisco's case he was working with data on the travel habits of hundreds of commuters whose point of origin was the northwest suburb of Skokie, and whose destination was the central

business district of Chicago, known as “the Loop”. One alternative for the great bulk of these commuters was to drive to work. This might have taken, for a given person, 75 minutes, including walking time from the parking lot to the place of work. The most natural alternative for that person would been to go by train. This would entail walking (or going by bus) to the train station, waiting for the train, taking the Skokie Swift Express from Skokie to Evanston, changing there to the Chicago El (for elevated) train; obviously involving some transfer time, then taking the El into the loop, and walking (presumably) to the workplace. Let’s say that for our given person all this involved 115 minutes of time, on average.

Hence we have the fact that going by car would save this person around 40 minutes, each way, per day. Suppose now that going by car entailed (at the time) an average monetary cost (including parking) of $16 per day, while going by train cost $8 per day. This would reveal a tradeoff of 80 minutes against $8, which means a time cost of $6 per hour. The inference is that the commuters facing this tradeoff who choose to go by train value their travel time at less than $6 per hour, and that those who choose to go by car value their commuting time at more than $6 per hour.

Lisco had data on hundreds of Skokie commuters, facing different tradeoffs, mainly because of the distances of home and workplace from the respective train stations. The data confirmed one’s common-sense expectations -- people with low incomes tended to choose the train, those with high incomes tended to go by car. But some low-income people nonetheless went by car; they tended to be those who lived far from the train station. Similarly, the richer people who nonetheless took the train tended to be those who lived close to the station.

One could probably reach a pretty good judgment on a relevant value of travel time, simply by grouping commuters by income level and forming histograms of the time/money

tradeoffs they accepted or rejected. Lisco's procedure was more precise than this, and made fuller use of the data. He used an econometric technique known as probit analysis. Probit analysis deals with the probability, in this case, of a given choice of mode. Using this device, Lisco was able to derive the implicit tradeoff between time and money that would lead people of given income levels to split 50-50 on their choice of mode. That tradeoff would then be the value of commuter time assigned to that income level.[2]

The resulting values of commuter time were far below the average hourly earnings of the commuter income groups. This result has also emerged from every similar study, and has become an accepted part of transport economics.[3]

In developing countries it is hard to find the kinds of data that one needs for careful econometric studies. However, two important facts can help surmount this obstacle. First is the fact that in developing countries we see that on-the-job drivers (of trucks, buses, taxis, etc.) represent a higher share of the traffic than in more advanced countries. For these one not only can but directly should use their hourly earnings as the value of their time. Second, we can consider that car owners typically have considerably higher income than truck or bus drivers. We combine this with the result of modal choice studies (that travel time for these people should be valued at only a fraction of their hourly earnings) to simply take the hourly earnings of truck

---

2

If P_c is the probability of choosing to go by car, X_c is the extra cost per hour saved as a result of going by car, and Y is the level of family income, one can fit a probit regression of the form $P_c = a - bX_c + cY$. Then, setting $P_t = 0.5$, one can solve for $X_c = (a-0.5+cY)/b$. This shows how X_c, the implicit value of commuter time, varies by income level.

3

One can make a personal check in this proposition in any place where commuters all tend to work in a given compact area, like the Loop in Chicago. What we find in such cases is that you can get all-day parking at lower rates, the farther you go from that central area. Thus, if you can save $1 by walking two blocks more, each way, and the average total walking time for four blocks is 10 minutes, then walking time is worth more than $6 an hour for those who park nearer, and less than $6 an hour for those who park farther away.

and bus drivers as also representing the cost of travel time to car owners. Not an elegant solution, perhaps, but a workable one.

Congestion Externalities

Even though the fraction of families owning cars is very much smaller in developing countries than it is in the developed world, the problem of traffic congestion seems pervasive at all levels. This is important because of the specific role that congestion externalities play in the economics of transportation.

The first lesson on this subject is that congestion externalities are present, even when we may not be conscious of their existence. How to explain this? Consider the fact that on a standard well-built concrete road, traffic will flow at something like 60 to 70 miles per hour, when traffic is light. But we all have experienced traffic flows at 20 to 30 or 40 miles per hour on these very same roads. Not only that, but we know why we see these lower speeds -- more vehicles are using the road. Thus only two points of observation -- one with high average speed and low traffic volume, the other with low average speed and high traffic volume -- are enough to reveal the basic principle. More vehicles on the road mean slower average speed. Highway engineers have plotted this relationship, and found it to be a continuous curve -- traffic does not go unimpeded at 60 mph up to 500 vehicles per hour, and then drop suddenly to 30 mph as volume exceeds 500. No, the relationship between average speed and traffic volume is a smooth curve, in which speed starts to decline long before one perceives noticeable congestion.

This is important, because it isn't just on those frustrating occasions that we feel ourselves "stuck in traffic" that congestion externalities exist. In fact, they are present most of

the time that vehicles are using the roads.[4] What happens is that one extra swath of traffic brings average speed down from 60 to 58 mph, another swath brings it down from 58 to 55, etc. -- until finally, after more and more vehicles have been added to the flow, average speed gets down to 40 and 30 and 20 mph.

The externality connected with traffic flow is that each added vehicle contributes a tiny bit to this general slowing of traffic. This can be appreciated by simply thinking of what a continuous curve relating average speed to traffic volume really means. Just as a bucket is filled by a huge number of successive drops, each of which brings its own equal contribution, so it is with traffic. Here, starting with quite modest volumes of traffic, each added vehicle contributes its tiny bit to the slowing down of the rest. If an extra 100 vehicles end up slowing traffic from 60 to 58, we can infer that each one of these has slowed the whole mass of users by 2/100, or .02 miles per hour. If going at 60 mph, their time cost was 10 cents per mile, one extra vehicle now brings this cost to 10.02 cents per mile. In the end everybody -- the "old" drivers and the "added" drivers, ends up going 58 mph. The added drivers perceive this cost. But they do not perceive the cost they inflict on the 1000 or 2000 "old" drivers who, but for the 100 newcomers, would have sped ahead at 60 mph.

To put this externality in terms that are easy to understand, we can make use of a very convenient approximation, which necessarily gives a conservative estimate of the time externality. A key element in this rule of thumb is what we call the "unimpeded speed" of traffic

---

4

When I say most of the time I do not mean most of the hours of the year. Rather, I mean most of the time spent by vehicle occupants. I once did a study of congestion externalities for five American cities. Their data revealed that something like 3/4 of the total traffic going to and from the city occurred between 6 and10 a.m., and between 3 and 7 p.m., on weekdays. During these hours, for sure, added vehicles would lead to lower average speed.

on the road. This is the average speed of traffic on that road when traffic is very light. This might be 20 mph on a dirt road, 30 mph on a gravel road, 40 mph on a blacktop road, 50 mph on a 2-lane high-type concrete road, 60 mph on a 4-lane divided highway and 70 mph on a multilane freeway. Call this unimpeded speed "a", and let "s" be the actual speed that you observe at a given point in time. Our approximation would then say that the externality, at that point in time, would be (a-s)/s times the average time cost per vehicle mile.

To give an example, suppose we are dealing with an ordinary blacktop road, on which we know (or guess) that the unimpeded speed is 40 mph. We are going on a Sunday drive, and we notice that our average speed is about 30 (as we move with the traffic). Our formula says the externality is (a-s)/s; or (40-30)/30 times our average time cost. If the vehicle hour is worth $7.50, then the vehicle mile is worth 25 cents (= $7.50/30 mph). So the externality would be 8 1/3 cents per mile that we are driving in our car. If the traffic were going only at 25 mph the externality would be (40-25)/25 times average cost. That is, 60 percent of $7.50/25, or (.6)(.30) or 18 cents per mile that we drive our car. This is a calculation that each of us can make, as we drive on different types of road and with varying densities of traffic. It can be quite instructive, and certainly helps to hone one's intuition about highway economics.

Let us go back to the project of improving a given road -- call it road H. Volume of traffic on H will increase as a result of the project, but we have already seen how the direct benefits on road H should be counted. We want now to look into what happens on other roads; call them A and F, Road A is an alternative (substitute) for road H (our project road). So when H is improved, part of the increment of traffic that we observe on H will have come from A. Thus road A gets to be less congested as a consequence of our project on road H. The gain on road A is measured by its own (a-s)/s, times its own time cost for the traffic diverted to H.

Road F is a feeder road into H, hence when H is improved, traffic on F increases. F becomes more congested and there is an external loss on that road, equal to its (a-s)/s times its average time cost times its increase in traffic volume.

On Critical Traffic Volumes and Stage Construction

We have seen how the benefits of a highway improvement project are linked to the volume of traffic that already prevails on that road and that can be expected to prevail, year by year in the future, in the absence of the project. If we start with a dirt road and ask when it should be turned into a gravel road, we have a pretty good idea of how much such an improvement would "typically" cost. Based on this typical cost we can make a technically founded guess as to how large a volume of existing traffic it would take, in order plausibly to justify the upgrading to gravel.

So it is with each stage. It may take only 200 vehicles a day to plausibly warrant shifting from dirt to gravel, but possibly 500 vehicles a day to make a reasonable case for converting a gravel road to blacktop. No one would argue that such numbers should be used to justify a particular gravel road or a particular blacktop project. But where they can be especially useful is in the pre-selection of cases to be studied in more detail. It is extremely easy to get a traffic count of how many vehicles pass given points on a road. The old technology of doing this by pneumatic tubes laid across the road is still perfectly functional, and very cheap. Newer technology of electronic sensors might turn out to be more convenient, and maybe even cheaper.

The idea is to use traffic counts as signals of which existing roads are good candidates for upgrading -- from dirt to gravel, from gravel to blacktop, from blacktop to concrete, from two lanes to three or four, etc. For each stage there would be a critical volume of traffic that would

cause certain stretches of road to be selected for more serious evaluation. I'm told that the French highway authorities have been doing this for many decades, to very good effect.

A close cousin of stage construction is varying standards for different segments or stretches of the same road. If one travels highway 101 between Los Angeles and San Francisco, one finds that in some stretches it has four lanes in each direction, in other stretches three, in still other stretches, only two. And if my memory serves me well, there are still other stretches in which it is a standard undivided highway of two or three lanes. This is the kind of variation that will quite naturally emerge from a system of serious economic project evaluation. All it takes to generate such a result is varying volumes of traffic as between one stretch and another. In the case of highway 101, traffic is extremely dense in the vicinities of Los Angeles and San Francisco. Volumes are much lower as we get 100 miles or more away from either of these centers. Hence it is perfectly sensible, from an economic point of view to have different construction standards for different stretches.[5]

---

[5] Highway engineers sometimes press for absolutely uniform standards along a road. Their argument is soundest when we think of stretches of just a mile or two in length. Here, changing from two lanes to four lanes, then back to three, etc., could invite serious costs connected with auto accidents produced as drivers are "surprised" by a sudden change in pavement or in number of lanes. But that same argument has little or no merit when one considers stretches of, say, ten or twenty miles or more. Here transitions can be clearly signaled well in advance. There need be no surprises!