Optimal Highway Capacity

Philip A. Viton

January 31, 2014
The Capacity Problem

Up to now we have implicitly taken highway capacity as given, and derived the associated congestion tolls as a function of the traffic on the road. And even though we have argued that road-building will not solve the congestion problem, it still poses some interesting questions:

- Is our current highway system over- or under-built?
- If we were planning a new highway, how big (wide) should it be?

We shall attempt to provide answers to these questions. Specifically, we would like to know:

- How much highway capacity should we provide?
- How much traffic should there be on that road (as opposed to the rest of the network)?
- What are the congestion tolls that would result in that traffic?
We consider building a limited-access highway between two points.
We measure road capacity by standard-sized (12’') lanes in the given direction. For now, we ignore issues of indivisibility of lanes.
For simplicity, we consider the length of the highway to be 1 mile.
Before the road is built, commuters use their best alternative route/mode, and incur the alternative costs.
In any period, each commuter makes one trip, just as in our previous discussion.
As before, we order commuters in any period by their next-best costs.
We work with annualized costs. That is, we convert items like construction costs, which are incurred up-front, to an equivalent cost that we imagine is being incurred every year. This is something like taking a present value and spreading it out over some fixed time.

For simplicity, we assume that total traffic volume does not change from year to year.

But we will allow traffic to vary over a year: for example, peak periods and other periods.
Suppose we build a new freeway. Some commuters will switch to it. What are the consequences?

1. The new freeway users no longer incur the costs of their previous mode. Call this savings their Gross User Benefits.

2. The new freeway users incur costs of using the freeway. These are the product of average freeway user costs and the number of freeway users. Call these the User Costs.


4. Finally, we need to take account of the (public-sector) costs of providing and maintaining the new freeway. Call these the Public Costs.
It therefore makes sense to think of choosing capacity (= the number of standard-sized lanes) to maximize a measure of Net Social Welfare, defined as:

\[ \text{NSW} = \frac{\text{Gross User Benefits} - \text{User Costs}}{\text{Public Costs}} \]

(of course these will need to aggregated over time).

This is the Optimal Capacity problem we’d like to solve.
If individual $i$ is assigned to the freeway, then his/her gain ($= \text{gross benefit}$) is the saving of the cost of the alternative, ie $1 \text{ trip } \times c_{ai} (= \text{shaded strip})$.

Similarly for any other individual, like individual $j$ in the figure.
We add up all the gross benefit strips.

So Gross User Benefits of assigning the first \(q^*\) users to the freeway is the area under the demand function from \(q = 0\) to \(q = q^*\).
To obtain their gross user benefits, freeway users incur an average user cost $AC(q^*)$.

Total freeway user costs are $q^* \cdot AC(q^*)$ (hatched area).

Net user benefits are gross user benefits minus user costs.

This is the area under demand function above cost ($= AC(q^*)$).

This is known as Consumer’s Surplus. In the figure, it is shaded and un-hatched.
Implementation

We now implement this formally.

- $w =$ standard-sized lanes in the given direction.
- $q_t =$ PCE’s in period $t$.
- $P_t(q_t) =$ (inverse) demand curve for travel in period $t$.
- $AC_t(q_t, w) =$ average user costs in period $t$. We assume that:

$$\frac{\partial AC_t}{\partial q_t} > 0 \quad \frac{\partial AC_t}{\partial w} < 0$$

We can think of $AC_t(q_t, w)$ as the time (cost) to drive 1 mile on a $w$-lane road when traffic in period $t$ is $q_t$.

- $\rho(w) =$ annualized public cost of providing 1 mile of $w$-lane road.
Model Formulation

- Gross Benefits to Users = area under demand function.
- Total User Cost = [Average User Cost] × [Number of Users].
- We also incur annualized public-sector (road provision) costs, $\rho(w)$ for the $w$-lane road.
- So Net Social Welfare is:

$$\text{NSW} = \sum_t \left( \int_0^{q_t} P_t(s_t) \, ds_t - q_t AC_t(q_t, w) \right) - \rho(w)$$

We want to maximize this in terms by choosing freeway traffic $q_t$ and road capacity $w$. 
The first-order conditions (FOCs) for a maximum are that the derivatives of NSW with respect to $q_t$ and $w$ are zero:

$$\frac{\partial NSW}{\partial q_t} = 0 : P_t - AC_t - q_t \frac{\partial AC_t}{\partial q_t} = 0$$

$$\frac{\partial NSW}{\partial w} = 0 : -\sum_t q_t \frac{\partial AC_t}{\partial w} - \rho'(w) = 0$$

(we will assume that the second-order conditions for a maximum hold).
Interpretation: First FOC (I)

Let’s think about these, beginning with the first FOC, which we obtained using Leibniz’ Law for differentiating under an integral.

- We have:

\[ 0 = P_t - AC_t - q_t \frac{\partial AC_t}{\partial q_t} \]

- Re-write it as:

\[ P_t = AC_t + q_t \frac{\partial AC_t}{\partial q_t} \]

- We have seen this before. It says that the price (cost) users should face in period \( t \) has two components: \( AC_t \) and \( q_t(\partial AC_t / \partial q) \).
Look at the difference between price $P_t$ (the cost commuters should face) and $AC_t$ (the average cost they do face):

$$P_t - AC_t = q_t \frac{\partial AC_t}{\partial q_t}$$

- $(\partial AC/\partial q_t)$: the average damage (harm) done by an additional PCE in period $t$.
- $q_t \cdot (\partial AC/\partial q_t)$: total harm done to the traffic stream by an additional PCE in period $t$. 
As we’ve seen, the congestion problem is that users decide on trip-making based on the average user cost incurred \((AC_t)\) and not on the marginal cost of their actions.

So our result says that for users in period \(t\) to face the right price (incentive) to make trips, we need to ensure that they also face a term

\[
\tau_t = q_t \frac{\partial AC_t}{\partial q_t}
\]

This will be the optimal congestion toll in period \(t\).

This was exactly our earlier result for the congestion toll, when the road capacity was not considered variable.
Interpretation: Second FOC (I)

Now look at the second FOC:

\[- \sum_t q_t \frac{\partial AC_t}{\partial w} - \rho'(w) = 0 \text{ or} \]
\[- \sum_t q_t \frac{\partial AC_t}{\partial w} = \rho'(w)\]

- \(\rho'(w)\) = annual marginal cost of providing additional lane capacity.
- \(- (\partial AC_t/\partial w) = \text{minus} \) the marginal average user cost = marginal user benefit, in period \(t\) of providing an additional lane.
- \([- q_t \cdot (\partial AC_t/\partial w)] = \text{total user marginal benefit of providing additional capacity in period } t.\)
- We add this quantity up over all periods \(t\).
Interpretation: Second FOC (II)

- So our condition amounts to: choose \( w \) so that

\[
\text{Marginal User Benefits} = \text{Marginal Public Costs}
\]

- To understand this, suppose we start with a very small number of lanes \( (w) \). Should we add an additional lane?
- Our answer asks us to consider on the one hand the benefits that this new lane would provide to users. These are the marginal user benefits.
- On the other hand it asks us to look at the costs of providing the new lane. These are the marginal public costs.
- Our criterion is: continue to add lanes as long as the marginal user benefits exceed the marginal public costs.
Consider the problem: choose $w$ to minimize Total Freeway Costs = Total Freeway User Costs + Total Public Costs:

$$\sum_t q_t AC_t(q_t, w) + \rho(w)$$

You can verify that the FOC for this problem is exactly the same as FOC$_2$ for the problem of maximizing Net Social Welfare.

This means that we can find the optimal capacity on the basis of costs alone (we don’t need to worry about demand functions).

And once we know the correct capacity, we can find the optimal tolls using our previous results.
As we stressed in the course Introduction, paying for roads is an important consideration for transportation planners.

Suppose we provide the optimal amount of capacity and implement the optimal congestion tolls, so a user in period $t$ pays a toll of

$$
\tau_t = q_t \frac{\partial AC_t}{\partial q_t}
$$

per PCE-mile.

What is the relation between total toll collections and total public costs?
Implications for Highway Finance (II)

Before we can answer this, we need a couple of technical results.

- First, suppose we increase both traffic and capacity by the same proportion (say, double both traffic and capacity).
- We will assume that when we do this, average user costs $AC_t$ are unchanged.
- This is plausible since as we’ve seen, speeds depend on the volume-to-capacity ratio, not on volume and capacity separately.
- This assumption has a technical name: we are assuming that $AC_t(q_t, w)$ is homogeneous of degree zero in $q_t$ and $w$.
- It turns out that if $AC_t(q_t, w)$ is homogeneous of degree 0 then, by a result known as Euler’s Theorem:

$$q_t \frac{\partial AC_t}{\partial q_t} + w \frac{\partial AC_t}{\partial w} = 0$$
Average and Marginal Construction Costs

The second technical result relates construction costs and returns to scale,

Claim: we have

\[ MC(w) = AC(w) + w \frac{dAC(w)}{dw} \]

Note that this is exactly the same result we had in our derivation of the congestion toll, where we phrased it in terms of average and marginal user costs. But if you want a proof: write the right-hand side using our notation for construction costs, as

\[ \frac{\rho(w)}{w} + w \frac{d}{dw} \left( \frac{\rho(w)}{w} \right) \]

Now use the quotient rule from calculus to expand the last term; you will find that the entire expression simplifies to \( \rho'(w) \), which is the marginal construction cost.
Claim: under constant returns to scale (CRTS) we have, equivalently:

\[
\begin{align*}
AC(w) &= MC(w) \\
\frac{\rho(w)}{w} &= \rho'(w) \\
\rho(w) &= w\rho'(w)
\end{align*}
\]
Costs and Returns to Scale (II)

To see this:

- Suppose we are using inputs $z$ to build $w$ lanes of road at a total cost of $\rho(w)$.
- Now suppose we double all inputs at fixed input prices.
- Clearly, total cost exactly doubles. But what happens to average cost?
- That depends on how much more output we get when we double all inputs.
- Under CRTS, by definition, doubling inputs results in output (lanes) exactly doubling. (Or, more generally, scaling up all inputs by the same proportion results in output increasing by exactly that proportion).
So under CRTS, average cost \((AC(w))\) is unchanged.

Hence \(dAC(w)/dw = 0\) and, using the results from slide 21, \(AC = MC\), and therefore

\[ \rho(w) = w\rho'(w) \]

as claimed.
Suppose instead that we are producing road capacity under decreasing returns to scale (DRTS).

Then doubling all inputs results in lane output increasing by less than a factor of two (output increases less than proportionately).

So average cost rises and \( \rho(w) < w\rho(w) \).

If there are increasing returns to scale (IRTS), then doubling all inputs results in \( w \) more than doubling (output increases more than proportionately).

So average cost decreases, and hence \( \rho(w) > w\rho(w) \).
Now let’s return to the tolls versus costs question.

- Toll in period $t$:

\[ \tau_t = q_t \frac{\partial AC_t}{\partial q_t} \]
\[ = -w \frac{\partial AC_t}{\partial w} \]

(using Euler’s Theorem).

- Now multiply this by the volume in period $t$ ($= q_t$) and sum over all periods $t$: the result is an expression for total annual toll collections.

\[ \sum_t \tau_t q_t = w \times \left( -\sum_t \frac{\partial AC_t}{\partial w} q_t \right) \]
With optimal investment we know (from FOC$_2$) that:

$$- \sum_{t} q_t \frac{\partial AC_t}{\partial w} = \rho'(w)$$

So:

$$\sum_{t} \rho_t \tau_t = w\rho'(w)$$
Now suppose that road construction is subject to constant returns to scale. Then marginal (public) cost = average (public) cost, so $\rho'(w) = \frac{\rho(w)}{w}$.

So if road building is characterized by CRTS:

$$w \rho'(w) = w \times \frac{\rho(w)}{w} = \rho(w)$$

And then:

$$\sum_t p_t \tau_t = \rho(w)$$

total toll collections = total road costs
So we have shown that

Under CRTS, toll collections will exactly cover the cost of the road.

Conversely, if there are increasing returns to scale (IRTS), toll collections will fall short of total costs, and the highway sector will need to be subsidized.

Under decreasing returns (DRTS), the road sector will have a surplus (profit).
Problem Setup (I)

We now turn to the problem of obtaining empirical answers. We turn first to average user costs.

- $AC_t(q_t, w)$ is the part of average user cost that depends on traffic and capacity provided.
- We ignore for now issues like pollution costs and fuel consumption.
- Then we may take $AC_t$ to be the user’s (average) time cost: the user’s valuation of the time spent traversing the 1-mile stretch of road.
- Write this as:

$$AC_t(q_t, w) = v \times \text{Travel time}$$

$$= \frac{v}{u_t(q_t, w)}$$

where $u_t$ is the space-mean speed in period $t$ and $v$ is the users’ value-of-time.
- This is the quantity we wish to minimize, via a choice of $q_t$ and $w$. 
As we have seen (slide 18) we can obtain the optimal lane capacity \( w \) by minimizing annualized Total Freeway Costs = Total Freeway User Costs + Total Public Costs.

Assuming that the value of time \( v \) is constant over time and the same for all users, we need to choose capacity \( w \) to minimize:

\[
TC = \sum_t v \frac{q_t}{u_t(q_t, w)} + \rho(w)
\]
We turn now to the empirical details. We need to know:

- What is $u_t$ (the attainable speed function)?
- What is $\rho(w)$ (the road construction cost function)? And what about returns to scale?
- What is the pattern of traffic over time (the $q_t$’s)?
- What is $v$ (the value of time)?
Empirical Implementation — Speeds

- We have already developed a few options for studying $u_t$.
- For now, we’ll use Keeler and Small’s (KS) estimated speed-flow curve for the Eastshore Freeway in the San Francisco Bay Area (SFBA).
- Then as we’ve already seen:

$$u_t = 45.94 + \sqrt{471.22 - 0.26 \frac{q_t}{w}}$$

(assuming that the capacity per lane $c = 2000$ PCE’s per hour).
Empirical Implementation — Public Costs

- For road construction costs, Keeler and Small studied the construction costs of the California Division of Highways in the SFBA in the early 1970’s.
- We use their original (1970s) estimates, but we can easily update them for inflation (see appendix).
- We divide public costs into:
  - capital (construction) costs.
  - land acquisition costs.
  - road maintenance costs.
After some analysis/experimentation, KS settle on the following functional form:

\[ \ln(KLM) = a_1 CRS + a_2 CUC + a_3 FR + a_4 FSU + a_5 FC + a_6 \ln(w) \]

where:

- \( KLM \) = 1972 construction cost per lane-mile.
- \( CRS \) = fraction of road that is in a rural area.
- \( CUC \) = fraction of road that is an arterial within city limits.
- \( FR \) = fraction of road that is a rural freeway.
- \( FSU \) = fraction of road that is an urban/suburban freeway.
- \( FC \) = fraction of road that is central city (within the cities of Oakland or San Francisco).
Note that this is a linear-in-parameters function, so it can be estimated using linear regression.

For this function, it is very simple to derive an estimate of the degree of returns to scale:

- If $a_6 = 0$ then we have CRTS.
- If $a_6 < 0$ then we have IRTS.
- If $a_6 > 0$ then we have DRTS.
Keeler-Small results:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRS</td>
<td>$a_1$</td>
<td>11.609</td>
<td>32.33</td>
</tr>
<tr>
<td>CUC</td>
<td>$a_2$</td>
<td>12.767</td>
<td>21.39</td>
</tr>
<tr>
<td>FR</td>
<td>$a_3$</td>
<td>12.993</td>
<td>17.82</td>
</tr>
<tr>
<td>FSU</td>
<td>$a_4$</td>
<td>13.255</td>
<td>17.19</td>
</tr>
<tr>
<td>FC</td>
<td>$a_5$</td>
<td>1.1151</td>
<td>2.07</td>
</tr>
<tr>
<td>$\ln(w)$</td>
<td>$a_6$</td>
<td>0.0305</td>
<td>0.078</td>
</tr>
</tbody>
</table>

$R^2 = 0.52$

Test for CRTS: $H_0 : a_6 = 0$. We do not reject, even though the point estimate indicates slight DRTS.
KS estimate land acquisition costs as a proportion of construction costs.
Dependent variable: Land acquisition costs/lane-mile ÷ Construction costs/lane-mile.
Results:

<table>
<thead>
<tr>
<th>Indep. Var</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRS</td>
<td>0.267</td>
<td>38.1</td>
</tr>
<tr>
<td>CUC</td>
<td>0.342</td>
<td>31.1</td>
</tr>
<tr>
<td>FR</td>
<td>0.300</td>
<td>14.3</td>
</tr>
<tr>
<td>FSU</td>
<td>0.323</td>
<td>32.3</td>
</tr>
<tr>
<td>FC</td>
<td>0.367</td>
<td>15.3</td>
</tr>
</tbody>
</table>
For maintenance costs, KS estimate: (t-statistics)

\[
\frac{\text{Maint Cost}}{\text{Lane-Mile}} = 2917 + 0.00045 \left( \frac{q}{w} \right) \\
(6.4) \quad (4.5)
\]

So the basic (= fixed) annual maintenance cost per lane mile is $2917.
Assume that we are interested in an autos-only highway.

KS estimate that construction costs would be 30% less for such a road than for the values already estimated, which were for a mixed-use road.

Assume that road lifetime is 35 years.

Assume that for discounting the appropriate rate is 6%.

We consider 3 road settings:

- Urban Central City Freeway: $FSU = 1$, $FC = 1$, all others 0.
- Urban-Suburban Freeway: $FSU = 1$, all others 0.
- Rural Freeway: $FR = 1$, all others 0.

Comments: we will be discussing the 35-year life assumption later (it is almost certainly too long). Today the 6% discount rate is probably too high; but we can easily experiment with different parameters.
### Empirical Implementation — Example Public Costs (II)

<table>
<thead>
<tr>
<th>Cost Component</th>
<th>Urban CC Freeway</th>
<th>Urb-Suburb Freeway</th>
<th>Rural Freeway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction Cost/Lane Mile</td>
<td>1,648,427</td>
<td>540,545</td>
<td>415,955</td>
</tr>
<tr>
<td>Portion for autos-only road</td>
<td>1,269,289</td>
<td>416,219</td>
<td>320,285</td>
</tr>
<tr>
<td>Annualized construction cost / lane-mile</td>
<td>86,767</td>
<td>28,456</td>
<td>21,898</td>
</tr>
<tr>
<td>Land Acquisition cost</td>
<td>465,829</td>
<td>134,439</td>
<td>124,787</td>
</tr>
<tr>
<td>Annualized land acquisition cost</td>
<td>27,950</td>
<td>8,066</td>
<td>7,487</td>
</tr>
<tr>
<td>Annual Maintenance Cost</td>
<td>2,917</td>
<td>2,917</td>
<td>2,917</td>
</tr>
<tr>
<td>Total Annual Cost per lane-mile</td>
<td>117,634</td>
<td>39,439</td>
<td>32,302</td>
</tr>
</tbody>
</table>

(All values in $$1972)
Empirical Implementation — Traffic Distribution

- We divide each day into 5 homogeneous periods.
- Total annual period- \( t \) traffic = \( q_t \times [\text{Hrs/Year for period } t] \).
- We express these as a fraction of major-direction peak traffic (called a peaking ratio), which is assumed to last 312 hrs/year.

<table>
<thead>
<tr>
<th>Period</th>
<th>Relative Traffic ( x_t )</th>
<th>Relative Duration ( n_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Peak, major direction</td>
<td>1.000</td>
<td>1.00</td>
</tr>
<tr>
<td>2. Near-peak, major direction</td>
<td>0.700</td>
<td>1.67</td>
</tr>
<tr>
<td>3. Day, major direction</td>
<td>0.500</td>
<td>7.08</td>
</tr>
<tr>
<td>4. Near-peak, day: minor dir.</td>
<td>0.383</td>
<td>5.41</td>
</tr>
<tr>
<td>5. Night, both directions</td>
<td>0.133</td>
<td>12.83</td>
</tr>
</tbody>
</table>
We follow Keeler and Small and consider two values of time: $1.50 and $3.00 per person-hour.

If we assume on average 1.5 people per car, this gives us estimates of \( v \) as $2.25 or $4.50 per PCE-mile.

Note that these are in $1972. See the appendix for an estimate of how they could be updated to today’s values.
The upshot of all this is that total annual user costs can be expressed as proportional to:

\[ TC = \sum_{t=1}^{5} \frac{v \cdot x_t \cdot n_t}{u_t} \]

Where now:

\[ u_t = 45.94 + \sqrt{471.22 - 0.26 \cdot x_t \cdot n_t \cdot x} \]

and \( x = \text{peak volume per lane} = \frac{q_1}{w} \) is the only unknown.

We minimize \( TC \) with respect to \( x \).
## Results

<table>
<thead>
<tr>
<th>Road</th>
<th>Peak, Major Direction</th>
<th>Near-Peak Maj. Dir</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$v$</td>
<td>$\mu$</td>
</tr>
<tr>
<td>Rural Freeway</td>
<td>$4.50$</td>
<td>1430</td>
</tr>
<tr>
<td></td>
<td>$2.25$</td>
<td>1680</td>
</tr>
<tr>
<td>Urb-Sub Freeway</td>
<td>$4.50$</td>
<td>1510</td>
</tr>
<tr>
<td></td>
<td>$2.25$</td>
<td>1730</td>
</tr>
<tr>
<td>CC Freeway</td>
<td>$4.50$</td>
<td>1780</td>
</tr>
<tr>
<td></td>
<td>$2.25$</td>
<td>1800</td>
</tr>
</tbody>
</table>

Interest rate: 6%; Flow in PCEs/lane-mile; speeds in mph; tolls in 1972 cents/PCE-mile.

Tolls are negligible in other periods.
Note that we characterize the optimum in terms of peak flow, not lanes.

But if we know the actual peak traffic (in PCEs per peak hour) we can compute the optimal lane capacity by dividing by the appropriate peak flow.

To deal with indivisibilities (lanes cannot be fractions) we can round to the nearest integer, or round both up and down, and check which solution has the lowest cost.

Note that doing this will change the “optimal” speeds and tolls, but these are easily re-computed.

Keeler and Small also consider the effect on user behavior of levying tolls by supposing that this would result in a less peaky distribution of traffic. See their paper for their results on this.
Theodore E. Keeler and Kenneth A. Small.
“Optimal peak-load pricing, investment and service levels on urban expressways”.
Appendix - Price Changes

- Rough estimate of price changes for road construction costs:
  - 1972 – 2013, third quarter: 7.22
  - Source: Caltrans

- General consumer price index (relevant to values of time):
  - 1972 – early 2014: 5.67
  - Source: Online Dollar Times Inflation Calculator