Optimal Highway Capacity

Philip A. Viton

January 19, 2012
Even though we have argued that road-building will not solve the congestion problem, it still poses an interesting question: is our current highway system over- or under-built? Here we shall attempt to provide an answer to that question.

Specifically, we would like to know:

- How much capacity should we provide on a given road?
- How much traffic should there be on that road?
- What are the congestion tolls associated with that capacity and traffic?
We consider a 1-mile stretch of road in a single direction.

We work in annualized quantities.

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

But we allow traffic to vary over a year: for example, peak periods and other periods.

We measure capacity by standard-sized lanes in the given direction.
We set this up as an optimization problem

We want to choose capacity (= standard-sized lanes) to maximize Net Social Welfare

Net Social Welfare = Net User Benefits – Public Costs

Net User Benefits = Gross User Benefits – User Costs

So

\[ \text{NSW} = [\text{Gross User Benefits} - \text{User Costs}] - \text{Public Costs} \]
If individual \( i \) is assigned to the freeway, then the gain ( = gross benefits) is 1 trip \( \times c_{ai} \) ( = shaded strip)\n
Similarly for any other individual \( j \)
We add up all the gross benefit strips.

So Gross User Benefits of assigning $Q^*$ users to the freeway is the area under the demand function from $Q = 0$ to $Q = Q^*$. 

$$D$$

$Q$
Indifference curve and budget line between money ($M$) and travel $Q$ for an individual.

In equilibrium, slopes are equal so ($p =$ cost of travel)

$$\frac{\Delta M}{\Delta Q} = -p$$
Then $\Delta M = -p \cdot \Delta Q$ : if we take away 1 unit of travel ($\Delta Q = -1$), required money compensation is $\Delta M$.

This is the shaded box in the demand curve picture: area $= p \cdot \Delta Q$.

Now take away one unit more and compensate with income (another little strip) ; do this until all $Q$ is taken away.

Total required compensation (= gross user benefits) in either case is area under demand function.
To obtain gross user benefits, freeway users must pay the average user cost $AC(Q^*)$.

- Total expenditures are $Q^* \cdot AC(Q^*)$ (hatched area)
- Net user benefits are gross user benefits minus expenditures
- This is area under demand function above price ($= AC$)
- Also known as Consumer’s Surplus.
Implementation — Concepts

- $w =$ 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 $t$
- $C_t(Q_t, w) =$ average user costs in period $t$. We assume

$$\frac{\partial C_t}{\partial Q_t} > 0 \quad \frac{\partial C_t}{\partial w} < 0$$

We can think of $C_t$ as the time (cost) to drive 1 mile in period $t$
- $\rho(w) =$ annualized public cost of providing 1 mile of $w$-lane road
Model Formulation

- Benefits to Users = Area under demand function
- Total User Cost = [Average User Cost] × [Number of Users]
- Net Social Welfare:

\[
\text{NSW} = \sum_t \left( \int_0^{Q_t} P_t(s_t) ds_t - Q_t C_t(Q_t, w) \right) - \rho(w)
\]

We want to maximize this in terms of \( Q_t \) and \( w \)
First-Order Conditions (FOCs)

The first-order condition for an optimum are that the derivatives of NSW with respect to $Q_t$ and $w$ are zero:

$$\frac{\partial \text{NSW}}{\partial Q_t} = 0 : \quad P_t - C_t - Q_t \frac{\partial C_t}{\partial Q_t} = 0$$

$$\frac{\partial \text{NSW}}{\partial w} = 0 : \quad -\sum_t Q_t \frac{\partial C_t}{\partial w} - \rho'(w) = 0$$
Interpretation: First FOC I

- First FOC:

\[ 0 = P_t - C_t - Q_t \frac{\partial C_t}{\partial Q_t} \]

- Re-write it as:

\[ P_t = C_t + Q_t \frac{\partial C_t}{\partial Q_t} \]

- We have already seen this.
Difference between price and (average) user cost:

\[ P_t - C_t = Q_t \frac{\partial C_t}{\partial Q_t} \]

- \((\partial C / \partial Q_t)\) : the average damage (harm) done by an additional PCE
- \(Q_t \cdot (\partial C / \partial Q_t)\) : total harm done to the traffic stream by an additional PCE
As we’ve seen, the congestion problem is that users decide on trip-making based on their average user cost incurred ($C_t$) and not on the full 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

$$\tau_t = Q_t \frac{\partial C_t}{\partial Q_t}$$

This is the (Optimal) Congestion Toll in period $t$.
Interpretation: Second FOC I

- Second FOC:

\[-\sum_t Q_t \frac{\partial C_t}{\partial w} - \rho'(w) = 0\]

\[-\sum_t Q_t \frac{\partial C_t}{\partial w} = \rho'(w)\]

- \(\rho'(w)\) = annual marginal cost of providing an additional lane
- \(- (\partial C_t / \partial w)\) = \textit{minus} marginal average user cost (\(=\) marginal user benefit) in period \(t\) of providing an additional lane
- \(-Q_t \cdot (\partial C_t / \partial w)\) = total user marginal benefit of providing additional capacity in period \(t\)
- We sum this over all periods.
- So our condition amounts to: choose \(w\) so that

\[\text{Marginal User Benefit} = \text{Marginal Public Cost}\]
Note that we get the second FOC by solving the problem: choose $w$ to minimize
\[
\sum_t Q_t C_t(Q_t, w) + \rho(w)
\]

ie minimize Total Social Costs (ie Total User Costs + Total Public Costs)

This means that we can obtain optimal capacity on the basis of costs alone (we don’t need to worry about demand functions).
Implications for Highway Finances I

- Suppose we provide the optimal amount of capacity and implement the optimal congestion tolls
- A user in period $t$ pays a toll of
  \[ \tau_t = Q_t \frac{\partial C_t}{\partial Q_t} \]
- What is the relation between total toll collections and total public costs?
Suppose we increase traffic and capacity by the same proportion (say, double both traffic and capacity)

We assume that when we do this, average user costs $C_t$ 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 $C_t(Q_t, w)$ is *homogeneous of degree zero* in $Q_t$ and $w$.

There is a mathematical result here. If $C_t(Q_t, w)$ is homogeneous of degree 0 then

$$Q_t \frac{\partial C_t}{\partial Q_t} + w \frac{\partial C_t}{\partial w} = 0$$

(Euler’s Theorem)
Toll in period $t$:

$$\tau_t = Q_t \frac{\partial C_t}{\partial Q_t}$$

$$= -w \frac{\partial C_t}{\partial w}$$

(using Euler's Theorem)

Now multiply by volume in period $t$ ($= Q_t$) and sum over all periods $t$: the result is total annual toll collections.

$$\sum_t \tau_t Q_t = -w \sum_t \frac{\partial C_t}{\partial w} Q_t$$
With optimal investment we know (from FOC$_2$) that

$$- \sum_t Q_t \frac{\partial C_t}{\partial w} = \rho'(w)$$

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 $= \text{average public cost and } $\rho'(w) = \rho(w)/w$

So under CRTS

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

And then

$$\sum_{t} p_t \tau_t = \rho(w)$$
Then, if there are CRTS in road construction, we have shown that toll collections will exactly cover the cost of the road.

Conversely, if there are IRTS, toll collections will fall short of total costs.

Under DRTS, the road sector will have a surplus (profit).
Empirical Implementation I

- $C_t(Q_t, w)$ 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 $C_t$ 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

$$C_t(Q_t, w) = \nu \times \text{Travel time} = \frac{\nu}{u_t(Q_t, w)}$$

where $u_t$ is the attainable speed in period $t$ and $\nu$ is the users’ value-of-time.
Assuming that $\nu$ is constant over time and the same for all users, we need to choose capacity $w$ to minimize total system costs, which are

$$\sum_t \frac{\nu Q_t}{u_t(Q_t, w)} + \rho(w)$$
Empirical Implementation III

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 $u_t$
- For the moment, we’ll use Keeler and Small’s estimated speed-flow curve for the Eastshore Freeway in the SFBA
- Then as we’ve 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)
For the road construction cost, Keeler and Small studied the construction costs of the California Division of Highways for roads in the SFBA in the early 1970’s.

(We can update these results for inflation).

We divide public costs into

- capital (construction) costs
- land acquisition costs
- road maintenance costs
Empirical Implementation — Construction Costs I

- After some analysis/experimentation, KS settle on the following functional form:

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

- where:
  - \(KLM = 1972\) construction cost per lane mile
  - \(\text{CRS} = \) fraction of road that is in rural area
  - \(\text{CUC} = \) fraction of road that is arterial within city limits
  - \(\text{FR} = \) fraction of road that is a rural freeway
  - \(\text{FSU} = \) fraction of road that is urban/suburban freeway
  - \(\text{FC} = \) fraction of road that is urban central city (within Oakland/San Francisco)
Note that this is a linear-in-parameters function.

For this function, there is a very simple returns to scale interpretation:

- 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>

$R^2 = 0.52$

Returns to scale: $H_0 : a_6 = 0$. We do not reject, even though the point estimate indicates slight IRTS.
KS estimate a function for land acquisition costs as a proportion of Construction Costs

Dependent variable: Land acquisition costs/lane-mile $\div$ 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>
Implementation — Maintenance Costs

- Results: (t-statistics)

\[
\frac{\text{Maint Cost}}{\text{Lane-Mile}} = 2917 + 0.00045 \frac{Q}{w} 
\]

- 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
### 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 Cost per lane-mile</td>
<td>117,634</td>
<td>39,439</td>
<td>32,302</td>
</tr>
</tbody>
</table>
We divide each day into 5 homogeneous periods

Annual traffic in period \( t = Q_t \times \text{Hrs/Year} \) for period \( t \)

We express these in terms of major-direction peak traffic, \( Q_1 \) lasting 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>
The upshot of all this is that the user costs can be expressed as

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

Where now

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

and \( x = \text{peak volume-per lane, } Q_1/w \)

We minimize user costs + public costs with respect to \( x \)
### Results

<table>
<thead>
<tr>
<th>Road</th>
<th>$v$</th>
<th>Peak, Major Direction</th>
<th>$\tau$</th>
<th>Near-Peak Maj. Dir</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Freeway</td>
<td>$4.50$</td>
<td>1430 55.9 2.7</td>
<td></td>
<td>60.5 1.1</td>
</tr>
<tr>
<td></td>
<td>$2.25$</td>
<td>1680 51.8 3.1</td>
<td></td>
<td>61.9 0.8</td>
</tr>
<tr>
<td>Urb-Sub Freeway</td>
<td>$4.50$</td>
<td>1510 54.9 3.3</td>
<td></td>
<td>60.0 1.2</td>
</tr>
<tr>
<td></td>
<td>$2.25$</td>
<td>1730 50.6 4.3</td>
<td></td>
<td>58.5 0.8</td>
</tr>
<tr>
<td>CC Freeway</td>
<td>$4.50$</td>
<td>1780 48.9 15.2</td>
<td>58.1 1.8</td>
<td>58.1 0.9</td>
</tr>
<tr>
<td></td>
<td>$2.25$</td>
<td>1800 47.7 58.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Tolls are negligible in other periods
When you’re not optimally providing capacity, tolls are straightforward: you simply supply the number of lanes on your road (and traffic and the relevant value of time).

No question of traffic distribution over time.
Freeway Tolls Without Construction I

- Setting: BPR speed-flow curve
- Free-flow speed 65mph; at-capacity speed 25 mph; lane capacity 2000 PCEs/hour
- Value of time: $4.50/hr in $1972. The effect of inflation is simply to multiply the tolls by the relevant inflationary factor
Freeway Tolls Without Construction II

Results (volume in PCEs per lane-mile; speed in mph; toll in dollars per PCE-mile):

<table>
<thead>
<tr>
<th>Volume</th>
<th>Speed</th>
<th>Toll/mi</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>64.60</td>
<td>0.00</td>
</tr>
<tr>
<td>1000</td>
<td>59.09</td>
<td>0.03</td>
</tr>
<tr>
<td>1500</td>
<td>43.15</td>
<td>0.14</td>
</tr>
<tr>
<td>1600</td>
<td>39.27</td>
<td>0.18</td>
</tr>
<tr>
<td>1700</td>
<td>35.42</td>
<td>0.23</td>
</tr>
<tr>
<td>1800</td>
<td>31.71</td>
<td>0.29</td>
</tr>
<tr>
<td>1900</td>
<td>28.22</td>
<td>0.36</td>
</tr>
<tr>
<td>1950</td>
<td>26.58</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Rough extrapolation: a toll of $0.40 per PCE-mile in 1972 would be about $2.13 per mile today.
Setting: BPR speed-flow curve

Free-flow speed: 40 mph; at-capacity speed 15 mph; lane capacity 1000 PCEs per hour

Value of time $4.50/hour in $$1972.

Note that arterial tolls involve a serious collection issue: it’s not clear how to implement this effectively, and whether it makes sense to toll just one arterial, and not the feeders, collector roads; or even the arterials that are substitutes for this one.
Arterial Tolls Without Construction II

Results (volume in PCEs per lane-mile; speed in mph; toll in dollars per PCE-mile):

<table>
<thead>
<tr>
<th>Volume</th>
<th>Speed</th>
<th>Toll/mi</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>39.79</td>
<td>0.00</td>
</tr>
<tr>
<td>250</td>
<td>38.02</td>
<td>0.01</td>
</tr>
<tr>
<td>500</td>
<td>30.90</td>
<td>0.08</td>
</tr>
<tr>
<td>750</td>
<td>22.08</td>
<td>0.23</td>
</tr>
<tr>
<td>800</td>
<td>20.47</td>
<td>0.27</td>
</tr>
<tr>
<td>900</td>
<td>17.54</td>
<td>0.36</td>
</tr>
<tr>
<td>950</td>
<td>16.22</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Rough extrapolation: a toll of $0.40 per PCE-mile in 1972 would be about $2.13 per mile today.
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 – 2011, third quarter : 7.2
  - Source: Caltrans

- General consumer price index (relevant to values of time):
  - 1972 – 2011 : 5.33
  - Source: Online Dollar Times Inflation Calculator
Appendix — Computer Programs

The 776 website contains GAMS programs for the models discussed here:

- `ks.zip`: models and data for Keeler-Small model

This model package contains parameters to allow you to adjust for inflation, and the degree of urbanization of the road you’re considering.