THE TRAFFIC INDUCEMENT EFFECT: ITS MEANING AND MEASUREMENT

Mark Hansen
University of California, Berkeley

INTRODUCTION

The "traffic inducement effect" of road improvements in urban areas is the subject of continuing controversy. Whether, to what extent, and under what conditions adding road capacity engenders traffic growth are, on the surface, empirical questions. However, these questions, like those of whether the death penalty deters crime, or whether welfare programs encourage teen pregnancy, have strong ideological overtones. These derive in part from the salience of the question to fundamental, highly contentious, issues in highway policy. The ideological dimension is enhanced because definitive answers to these questions have proved illusive, a consequence of our inability to conduct the relevant controlled experiments.

This paper seeks to contribute to the ongoing debate about the relationship between road supply and traffic in several ways. First, we reflect upon the policy context of the debate. Second, we seek to make the questions in dispute more precise by defining metrics that capture the impact of road supply on road traffic. Third, we report on research that has attempted to measure these impacts. Fourth, and finally, we offer recommendations for improving our ability to monitor and document on an ongoing basis how road improvements, and perhaps other transportation investments as well, influence traffic and travel in urban regions.

POLICY CONTEXT

In the 19th Century, as roads, canals, and railroads were built across the United States, analysts of the day distinguished between projects that were "developmental" and those and were "exploitive."[1] Developmental projects were expected to generate and serve new markets by enabling settlement of previously inaccessible hinterlands. Exploitive ones, in contrast, targeted existing markets, offering improved service compared with pre-existing alternatives. In a country where settlement of new territory was an urgent priority, projects in the former category were, as the above terms suggest, considered to have the higher purpose.

By their nature, developmental projects induced demand, whereas exploitive ones were more likely to divert pre-existing traffic. So in the 19th Century, traffic inducement was a considered a desirable impact of transportation improvements. In certain contexts, the same holds true today. For example, urban and intercity rail proponents stress (perhaps exaggerate) the ability of such systems to alter settlement and traffic patterns in ways that will stimulate traffic on the systems they advocate.

When it comes to roads, however, the tables are turned: advocates of road improvements view them a accommodations to largely exogenous demand, while opponents argue that such accommodation will inevitably spur more traffic in an endless spiral of road building and road filling.

Why this difference? In part, it has an economic interpretation, illustrated in Figure 1. Figure 1a depicts the impact of a road improvement, represented as a downward shift of the supply (average user cost versus traffic) curve for a road (or road network) from **s** to **s'**. If the demand curve (traffic level versus average user cost) is, like **d**, vertical--implying no induced traffic--the increase in consumer surplus resulting from the improvement is the rectangle **ABCD**. Conversely, if the demand curve is sloping like **d'**, so that some traffic is induced, the benefit is the smaller area **ABC'D'**. The difference derives from the fact that the supply curves are upward sloping i.e. that roads are subject to congestion effects.

In Figure 1b, the effect of an improvement to a transportation system not subject to congestion is shown. It can be seen that in this case, a sloping demand curve implies a greater benefit than a vertical one. The effect is even stronger when supply curves are downward sloping, due to economies or scale. Elastic demand also results in greater benefit if improvement and improvement of a congested system leads to an uncongested system at the new equilibrium (imagine that s' in Figure 1a remains flat beyond traffic level where it intersects with d'.)

Thus, a conventional welfare analysis offers on explanation for why induced traffic is seen in a negative light for road projects and a positive light for many others. In fact, however, such an analysis is ambiguous even in the case of a congestable road facility. But the negative view of induced traffic is also related to how we view the adjustments represented by movements along the demand curve from a normative standpoint. In the case of roads,

Figure 1a. Welfare Analysis of Transporation Improvement with Congestion

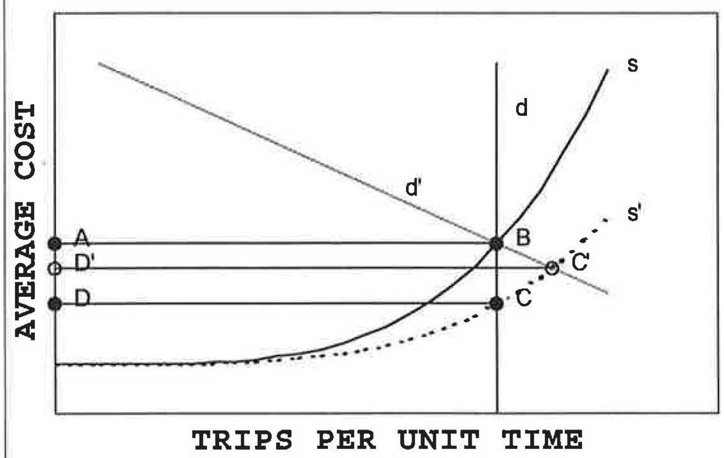
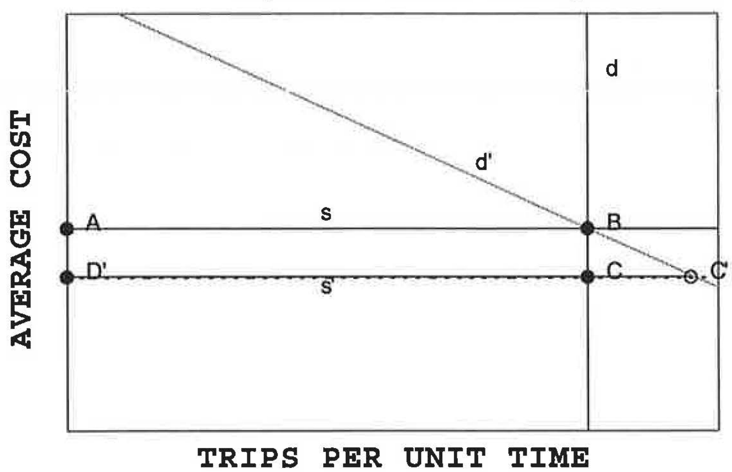


Figure 1b. Welfare Analysis of Transportation Improvement without Congestion



increased demand is associated with urban sprawl, increased fuel consumption, more emissions, and other ills associated with motorization. On the other hand, adjustments associated with increased rail use, such as more focussed development patterns or curtailed automobile use, are seen more favorably. Road advocates might, as they could in the welfare analysis, challenge these viewpoints, for example by pointing to increased road traffic as an indication of an invigorated economy, or of more households realizing the dream of owning a single family home in the suburbs. They have for the most part avoided this line, however, instead maintaining that the benefits or road improvements derive almost entirely from reduced congestion to an essentially fixed quantity of traffic.

MEASUREMENT ISSUES

The debate over the impact of road investments on traffic levels is sometimes caricatured as one over whether roads do or don't generate traffic. But this is not really the issue.

It is widely accepted that, holding traffic levels fixed, additional road capacity reduces traffic delays and travel times. It is also widely accepted that these impacts, by making vehicle travel more convenient, will tilt decisions about whether, where, and how to travel toward choices that involve increased vehicle usage. Neither of these generalizations is iron-clad. There are the familiar "paradoxes" in which a road improvement can lead to a redistribution of trips that result in increased travel times. There are also scenarios where a road improvement makes near-in places more accessible, altering activity and traffic patterns in a manner that reduces overall travel. (The results of Putman [2], suggest that this could occur if the Golden Gate Bridge were double-decked so that the San Francisco commuter shed shifted toward Marin and away from the East Bay.) Such counterexamples are, however, widely viewed as rare exceptions to the general rules. To the extent that the latter hold, it is a matter of logic that roads induce traffic to some degree.

The debate, therefore, is not over whether the effect exists, but its magnitude. This raises the question of metrics. Imagine that we have two identical regions with identical transportation systems and that, at some time t=0, we make a set of road capacity enhancements in one of the regions but not in the other, and that this is the only way in which we treat the regions differently. Over time one could monitor traffic levels in the two regions which, although equal at t=0, would presumably diverge thereafter as a result of the change to the road supply. Suppose we could characterize the magnitude of the road capacity change, as ΔS (or $\Delta \log(S)$), and the magnitude of the interregional traffic level difference a time t as $\Delta Q(t)$ (or $\Delta \log(Q(t))$). In this idealized situation, the traffic inducement effect of the capacity increase might be measured either as a simple ratio, $\Delta Q(t)/\Delta S$, or as an elasticity, $\Delta \log(Q(t))/\Delta \log(S)$. We prefer the latter, which we term the capacity elasticity of traffic, for two reasons. The elasticity indicates directly how a capacity increase affects the ratio of traffic to capacity, a widely accepted measure of level of service. Second, for a given elasticity and given capacity increase, the quantity of traffic induced varies directly with the ratio of traffic to capacity. This is plausible, since the higher ratio implies a higher level of congestion in the baseline situation.

The procedure in the above hypothetical experiment is not yet precisely defined, since we did not specify how S and Q are to be measured. Different procedures give different elasticities, each with its own significance. For example, if the capacity change involves the widening of a specific segment of road, than S could be the lane-width of that segment and Q the traffic on that segment. Alternatively, S and Q could be measured over larger subsets of the regional road networks, up to and including the networks in their entirety. When such aggregation is performed, it makes sense to measure S in terms of lane-miles and Q in terms of vehicle-miles.

With the measurement procedure specified, a given experiment like the one defined above would yield a specific set of measurement results--calculated capacity elasticities of traffic for different points in time after the road supply change in one of the regions. However, if the same experiment were performed using a different pair of regions, or using the same pair but a different road supply change, it is likely that different elasticity values would be obtained. These elasticities are not physical constants, but variables that depend in a complex way on characteristics of the region, its baseline transportation network, and the road supply change. But despite variation, the elasticities will have a central tendency, which could, in principle, be estimated by repeating the experiment using different regions and supply changes considered representative of the "populations" of interest.

TWO RECENT STUDIES

This section summarizes two recent studies whose objective was to estimate capacity elasticities of traffic. In one study [3], the elasticity was estimated at the road segment level. In this case, the question is: how does traffic on an individual road segment respond to an increase in capacity of that segment? In the second study [4], the elasticity is measured at the metropolitan area level. Here we are interested in how an increase in area-wide highway capacity affects area-wide highway traffic. Both studies focus on California metropolitan areas, and employ data for the last 2-3 decades.

The thought experiments described in the last section are useful for explaining what we are attempting to measure, but cannot actually be undertaken. Instead, we must devise quasi-experiments using real-world data. In both of segment-level and area-level studies, our quasi-experiments have used panel data. In the segment-level study, the panel consists of highway segments, while in the area-level study, it is metropolitan areas. In both studies, we follow the panel over time, using statistical methods to attempt to relate changes in traffic levels to changes in capacity levels.

Segment-Level Study

The panel consists of 18 highway segments, all of freeway or expressway grade, whose capacity was increased sometime in the late 1960s or 1970s. The segments are located in metropolitan areas–nine in the Los Angeles area, six in the Bay Area, two in Sacramento, and one in San Diego. The capacity expansions involved adding lanes--either one in each direction (11 cases), two in each direction (six cases), or a combination of one and two lane expansions (one case).

For each segment, annual average traffic count data, published by Caltrans, were obtained for selected years prior to and after completion of the capacity expansion. The years selected are those 1,4,7,10,... years before the capacity expansion and 1,4,7,10,... years after. The expansions themselves occur over a 1 to 4 year period over which no observations are included. Only years between 1960 and 1990 are included, so a given segment will have more observations prior to (after) the capacity expansion the later (earlier) the expansion took place. The maximum number of years after the expansion for which observations were available was 19, but only three segments have count data available for this time slice.

The data were used to estimate a model of the form:

$$\log(Q_{it}) = \alpha_i + \beta\cdot\log(C_{it}) + \gamma\cdot\log(SQ_t) + \lambda\cdot\frac{NC_{it}}{t^{\sigma}} + \epsilon_{it}$$

where:

- Q_{it} is the traffic volume of segment I in year t (t is measured from before the beginning or after the completion of the capacity expansion);
- C_{it} is the capacity (number of lanes) of segment I at time t;
- SQ_t is vehicle-miles traveled on the California state highway system in year t;
- NC_{it} is the ratio of capacity added to total capacity for t > 0, and zero for t < 0;
- $\alpha_i,\beta,\gamma,\lambda,\sigma$ are coefficients to be estimated;
- ϵ_{it} is a stochastic error term, drawn from a normal distribution with mean zero.

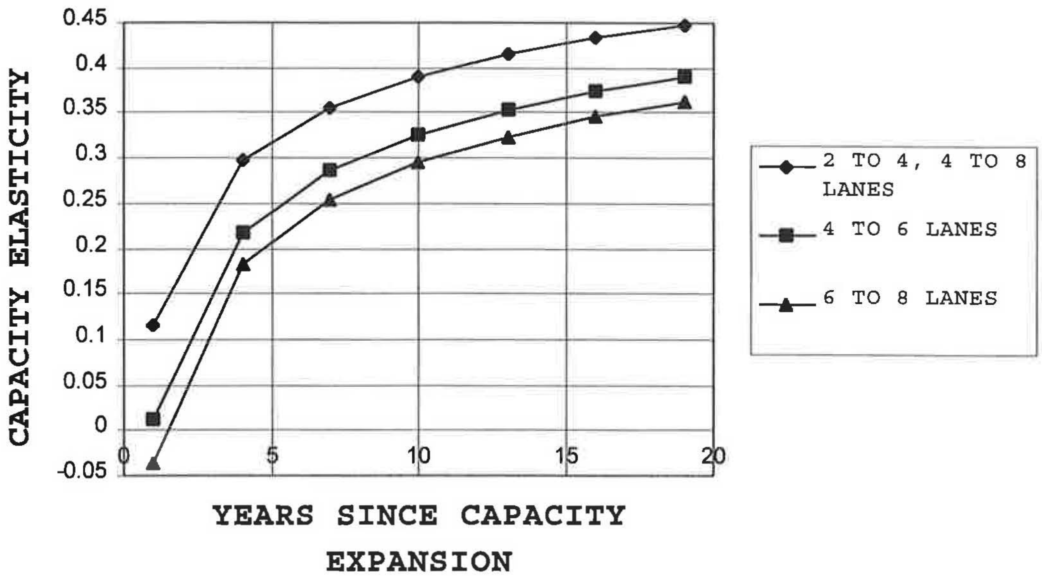
In this model, the hypothesis that traffic is unrelated to capacity implies that $\beta=0$ and $\lambda=0$. In that case, traffic on segment I in year t would be determined by the segment specific factor, α_i, and a time-specific factor related to the overall traffic level on California state highways, $\gamma\cdot\log(SQ_t)$. In other words, traffic on each segment would grow from a segment-specific baseline level, tracking growth of overall traffic on the state highway system. The hypothesis that traffic responds instantaneously to a change in capacity implies that $\beta>0$ and $\lambda=0$. In this case, an increase in capacity would immediately result in an upward shift of traffic, over and above any increase associated with statewide traffic growth. Finally, if $\lambda<0$, the response of traffic to new capacity is gradual. One year after the expansion, $\log(Q_{it})$ is $\lambda\cdot NC_{it}$ less than it would be if the new capacity were not new. As the time since the capacity expansion increases, this difference decreases with $t^{-\sigma}$.

This model was estimated on the panel data set described above. The model is linear in all coefficients except σ. In the estimation, we assumed different values for this coefficient, and used least squares to estimate the remaining coefficients, ultimately choosing the model with the σ value that yielded the best fit. The estimation results appear in Table 1. The β estimate is positive and significant, while the λ estimate is negative and significant, across the range of σ values that give the best fits. This implies a positive, non-instantaneous response of traffic level to an increase in capacity.

The estimation results can be used to calculate segment-level capacity elasticities of traffic for different times after the expansion. The elasticity will of course depend on the amount of time since the capacity expansion and the ratio of expanded to original capacity. The elasticities for the model with $\sigma=0.20$ are plotted against time since expansion, for different capacity increases, in Figure 2. The elasticities increase sharply during the first four years after the expansion, and more gradually thereafter. Four years after expansion, elasticities are in the 0.2-0.3 range. After 10 years, the elasticities are in the 0.3-0.4 range. Thereafter, increases are very slow, so that by 16 years after project completion the elasticities range from 0.35 to 0.43. Throughout, the highest elasticities are associated with larger fractional capacity expansions.

Throughout the period plotted, the elasticities are well below 1.0. This implies that the capacity expansion yields a sizable reduction in the ratio of traffic to capacity. In other words, although it appears that expanding the capacity of a highway segments results in an increase in segment traffic, there is still a substantial level-of-service gain.

Figure 2. Capacity Elasticity of Traffic, Road Segments



Area-Wide Study

In this study our panels consisted or urban areas, rather than highway segments. Our basic data consisted of state highway vehicle-miles traveled (VMT), state highway lane-miles, population, and per capita income for every urban county in the state of California, for the years 1973-1990. In one analysis, this panel was used directly. In a second analysis, the county-level data were aggregated to the metropolitan level--for example, observations from 10 counties considered by the federal government to belong to the San Francisco-Oakland-San Jose metropolitan area combined into one observation. As in the previous study, we sought to use this data to estimate a general relationship between road supply and traffic. Our basic model was:

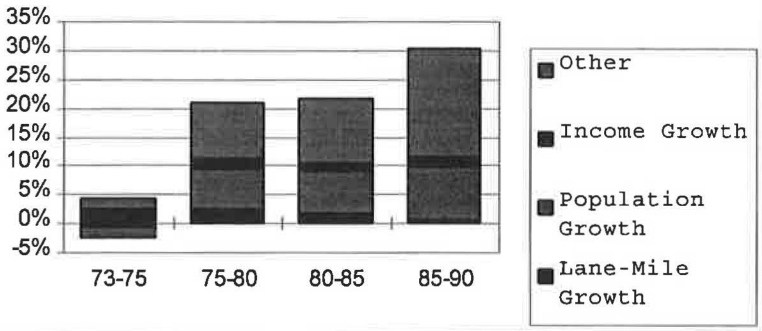
$$\log(VMT_{it}) = \alpha_i + \beta_t + \gamma \cdot POP_{it} + \psi \cdot PCI_{it} + \sum_{l=0}^{L} \omega^l LM_{it-l} + \epsilon_{it}$$

where:

- VMT_{it} is vehicle-miles traveled in area I and year t;
- POP_{it} population in area I and year t;
- PCI_{it} is income per capita in area I and year t;
- LM_{it-l} is state highway lane-miles in area I and year t-l;
- $\alpha_i, \beta_t, \gamma, \psi, \omega^l$ are coefficients to be estimated;
- ϵ_{it} is a random variable drawn from a normal distribution with mean zero.

This model contains fixed effects for both areas and years, the α_i and β_t respectively. In a world in which year to year changes in VMT were the same for each region, these fixed effects would explain all of its variation. If regions with higher population or income growth experience greater traffic growth, these effects will be captured by the γ and ψ coefficients. Finally, if after controlling for regional, time period, population, and income effects, covariation between road supply and VMT persists, this is captured by the ω^l coefficients. If the VMT response to a change in road supply were immediate, then only the ω^0 term would be positive. If coefficients on the lagged lane-miles variables are also positive, this implies that VMT response occurs over a period of time, with the complete adjustment occurring after L years.

Figure 3. Sources of California State Highway Urban VMT Growth, 1973-1990



To estimate these models, the value of L must first be determined. We did this by starting a L=0 and then incrementing L by 1 until we found the model that had the best statistical performance, labeling its associated L value as L*. Next we determined the appropriate number of free parameters to allow for the ω^l . At one extreme, we could allow each of these coefficients to vary arbitrarily, while at the other we could force them all to have the same value. As before, we sought the choice be-tween these extremes that yielded a model with the best statistical performance. This turned out to be the model that forced all the ω^l coefficients to have the same value.

Table 2 contains the estimated coefficients for the preferred models at both the county and metropolitan levels. L* is found to be two years for the county-level model and four years for the metropolitan level model. The estimated ω^l value, restricted to be the same for all l from 0 to L*, is 0.21 for the county model and 0.19 for the metropolitan model. In both cases, these estimates are highly significant statistically. To calculate the long-run capacity elasticity of traffic—the effect we expect to see L* years after a capacity expansion--we need only multiply ω^l by L*+1. The resulting elasticity is 0.62 in the case of the county-level model and 0.94 for the metropolitan model.

Table 2 also shows that population has a strong effect on traffic, yielding population elasticities of 0.46 at the county level and 0.69 at the metropolitan level. Per capita income, in contrast, has a small effect, particularly at the county level. It is interesting to calculate how VMT change if population and road supply grow by the same amount, so that road supply per capita remains constant. At the county level, 1 percent increase in both population and road supply will result in a 0.62+0.46, or 0.96, increase in traffic–not statistically different from 1 percent. At the metropolitan level, the same scenario yields a traffic growth of 0.69+0.94, or 1.63 percent. Presumably, the traffic-generating impact of growth at the metropolitan level is stronger because it involves increased intercounty travel, an effect not readily captured in a county-level analysis.

The estimation results can be used to estimate the contributions of population, income, road supply, and "other factors" to the overall growth in VMT that has occurred over the past two decades in California's urban regions. The first three effects are estimated using the estimated elasticities and the average growth in population, income, and road supply for California metropolitan regions. The effect of "other factors" is captured by trends in the time period fixed effects (β_t). The results at the metropolitan level are shown in Figure 3, which reveals that population growth has been the most consistent contributor to VMT growth over the past two decades. Since 1980, "other factors", presumably a combination of demographic, life-style, and gasoline price effects, have also played a major role. In contrast, the contribution of increased highway supply to VMT growth has been modest, particularly during the 1980s when the supply grew very slowly.

The above results all pertain to state highway VMT, rather than total VMT. In California, about 50 percent of total VMT is on state highways. A natural question is therefore whether the additional state highway VMT that seems to result from added lane-miles is diverted from local roads and streets. Data to definitively answer this question are, unfortunately, lacking. Published estimates for total county VMT are available only for selected years. Furthermore, these estimates are based on gasoline sales data rather than direct traffic counts, and are therefore not very reliable. Nonetheless, we used the data available to look for a relationship between state highway lane-miles and off-state highway VMT. If the diversion hypothesis is correct, than we would expect a negative relationship

TABLE 1 ESTIMATION RESULTS, SEGMENT TRAFFIC MODEL

COEFFICIENT (VARIABLE)	COEFFICIENT ESTIMATE, BY ASSUMED σ VALUE		
	σ = 0.05	σ = 0.20 [1]	σ = 0.75
β (Road Capacity)	1.30 (3.70)[2]	0.86 (4.70)	0.46 (4.53)
γ (Fraction of new road capacity)	-1.59 (-3.07)	-1.03 (-3.68)	-0.44 (-3.04)
λ (State Highway VMT)	1.06 (19.23)	0.96 (15.29)	0.96 (14.54)
Adjusted R^2	.9568	.9580	.9568

Notes
1. Preferred model, based on adjusted R^2.
2. t statistics in parentheses.

between these variables. Using a model of the same form as the one described above, we find a strong, positive, but statistically marginal, relationship between state highway lane-miles and off-state highway VMT at the county level, and a very weak, negative, statistically insignificant relationship at the metropolitan level. These results certainly do not support the diversion hypothesis; nor can they, given the limitations in the data, definitively refute it.

Discussion

Taken at face value, these results show a significant positive impact of road supply on traffic. Moreover, they suggest the impact, as measured by the traffic capacity elasticity, becomes stronger as the level of aggregation increases. At the road segment level, the long-run capacity elasticity of traffic is in the 0.3-0.4 range. The comparable figures at the county and metropolitan level are 0.6 and 0.9 respectively. This pattern implies that much of the traffic induced by a particular capacity expansion project occurs *away* from the expanded segment. It is a classic example of a network effect arising from complementarity between links: in order to avail themselves of the improved level of service on the expanded link, drivers used other links to access it. While level of service on the expanded link improves markedly, induced traffic on other links leads to marginal increases in congestion elsewhere in the system. We cannot, on the basis of our findings, assess the net impact of expanded capacity on the level of service provided by the road network. It interesting to note, however, that at the metropolitan level the long-run capacity elasticity of traffic is fairly close to 1.0, the value at which induced traffic is enough to maintain a constant ratio of VMT to lane-miles.

Our findings are less consistent with regard to the dynamics of the response to new capacity. The area-level findings suggest a response time of less than five years, while at the segment level there is evidence of continued adjustment 5, 10, or more years after the capacity is added. On the other hand, the latter results indicate that the response after five years is dramatically slower than that in the earlier years. Perhaps the longer term response is merely an artifact of the model employed in the segment-level analysis, or perhaps it is real but lost in the statistical noise of the area-level data.

There are important grounds for skepticism however. Perhaps the most important to concerns the direction of causality. Our analysis assumes that road supply is the cause and traffic level the effect. But one might argue that in fact the causality runs in the opposite direction, or in both directions. Thus, where we claim that traffic grew as a result of adding road capacity, others might counter that road capacity was added in response to, or anticipation of, this traffic growth, which would have occurred anyway.

Our use of panel data sets reduces the potential distortion arising from this problem of mutual causality. To see this, consider the road segment analysis. Suppose instead of the procedure we followed, we simply compared traffic volumes on highway segments with different capacities. Then, it would clearly be inappropriate to attribute the difference in traffic level entirely to the difference in capacity--almost certainly, the wider road is wider in part because it has to carry more

TABLE 2 ESTIMATION RESULTS, AREA TRAFFIC MODELS BY GEOGRAPHIC UNIT OF ANALYSIS

COEFFICIENT (VARIABLE)	COUNTY PANEL MODEL, P-W[1] ESTIMATE	METROPOLITAN PANEL MODEL, P-W ESTIMATE
γ (Population)	0.46 (9.03)[2]	0.69 (3.92)
ψ (Per Capita Income)	0.05 (0.88)	0.21 (1.87)
ω^l (Lane-Miles)[3]	0.21 (5.33)	0.19 (4.20)
R^2	0.994	0.997
L^*	2	4
Long-Run Capacity Elasticity of Traffic[4]	0.62	0.94
Number of Observations	480	196

Notes

1. Prais-Winsten estimates. This is a least squares technique that corrects for serial correlation in the data, see [4].
2. t statistics in parentheses.
3. Coefficient applies to current lane-miles and lane-miles 1,2,...,L^* years before.
4. The percentage increase in VMT resulting from a 1 percent increase in lane-miles, after a sufficient period of time for the full effect to be realized. Equal to ω^l coefficient times L^*+1, with any differences in table due to rounding.

traffic. But this is not what we did. Rather, we followed traffic levels on a number of segments, and found that traffic growth on these segments accelerated, compared to traffic growth on the state highway system as a whole, *after* capacity was added. One could still argue that highway planners, in their infinite wisdom, foresaw when this accelerated growth would occur and added capacity in anticipation of it. There is no statistical analysis that can refute such a claim, but one must question whether the processes of planning and delivering highway capacity expansions, lengthy, political, and fiscally constrained as they are, can be so responsive. An analogous argument holds at the area level.

CONCLUSIONS AND RECOMMENDATIONS

We have presented evidence that adding road capacity generates traffic. The effect is "strong" in the sense that the proportionate increase in traffic is of the same magnitude, although smaller than, the proportionate increase in capacity. The effect is stronger at the aggregate network level than at the individual link level. Most of the response occurs within five years of the capacity expansion. Although the effects observed may derive from diversion of traffic from local roads, the limited evidence available does not support this interpretation.

The findings are subject to several caveats. First, they apply to urban highways for a single state, California, over a limited time period, the 1970s and 1980s. Second, they are based on pooled data, and therefore do not reliably characterize the impacts of any specific capacity enhancement project, or road improvement program in a particular urban area. Third, they are not based on controlled experiments, but rather evidence gathered from quasi-experiments. As noted above, statistical correlation of quasi-experimental data cannot prove causality in a particular direction. Such an interpretation must rest upon one's a priori understanding of the processes at work.

Most importantly of all, our findings do not demonstrate that adding road capacity is a bad idea. While opponents of road construction had traditionally emphasized the phenomenon of induced demand, and road advocates de-emphasized it, it is not obvious that induced demand detracts from the social value of road improvements.

Much is to be gained from additional retrospective studies of the impact of road capacity enhancements, and other transportation investments, on traffic, travel, system performance, and economic welfare in urban areas. These efforts should be accompanied by more concerted attempts to incorporate the findings of retrospective studies into the methods and models used in traditional, future-oriented, planning activities. Despite the substantial effort that has gone into developing and refining such techniques, surprisingly little is known about their reliability and accuracy in predicting the consequences of transportation improvements. We must strive for convergence between results of analyses like those presented here and the detailed, predictive models necessary for planning. When such convergence has been achieved, the induced traffic debate can be laid to rest, and we can turn our argumentative energies back to welfare and the death penalty.

REFERENCES

1. Carter Goodrich, Government Investment in American Railroads and Canals: 1818-1878, Academic Press, 1960.

2. Stephen Putman, Integrated Urban Models, Pergamon Press, 1983.

3. Mark Hansen et al., The Air Quality Impacts of Urban Highway Capacity Expansion: Traffic Generation and Land-Use Change, UC Berkeley Institute of Transportation Studies Research Report 93-5, Chapter 3.

4. Mark Hansen and Yuanlin Huang, Road Supply and Traffic in California Urban Areas, forthcoming in Transportation Research A.