Induced Traffic: Review of the explanatory models

Philipp Fröhlich, ETH Zürich

Conference paper STRC 2003
Session Microsimulation

STRC | 3rd Swiss Transport Research Conference
Monte Verità / Ascona, March 19-21, 2003

Induced Traffic: Review of the explanatory models

Philipp Fröhlich
IVT
ETH Zürich
Zürich

Phone: +41-1 633 31 96
Fax: +41-1 633 30 57
e-Mail: froehlich@ivt.baug.ethz.ch

Abstract

A common claim in transport planning is that more road capacity generates additional car travel. While most authors argue that this additional traffic reflects an increase in the consumer surplus of the travellers, others see these additional trips and miles as wasteful.

The induced traffic demand from increased infrastructure supply over time typically coincides with an increase in population and a higher per capita income in the study area. To isolate the effects of the different variables is a complicated task. There is also a danger of oversimplification in answering this complex problem. This occurs quite frequently during discussions of traffic growth.

The present paper starts by exploring the concept of induced traffic, after which the major methodologies currently in use are analysed and discussed. The focus of the review is the various econometric models and approaches, as well as their results. Published findings reviewed here show an elasticity of daily vehicle mileage on lane miles for the short and long run of 0.2 – 0.3 and 0.6 – 0.8, respectively.

The paper concludes by suggesting central considerations for future work.

Keywords

Econometric Model – Induced Traffic – Elasticity – 3rd Swiss Transport Research Conference – STRC 2003 – Monte Verità

1. Introduction

The elasticity of demand with supply side changes is a basic economic concept. In the transportation world, we are also seeking equilibrium points; here, the demand meets the supply function. In highway planning, an expanded or newly built motorway in urban areas commonly nears its capacity limit considerably earlier than it was expected.

In the last 30 years, the interested public and transport professionals have discussed, and sometimes argued about, the cause of growth in transport demand. In the transport system, the supply side is treated as endogenous, whereas variables such as population, employment, income, or petrol price are considered exogenous.

In this respect, there are two primary opposing points of view. In the first, the growth of car traffic is primary influenced by the construction of new roads, mainly motorways, and this change on the supply side generates the additional traffic growth. Others, meanwhile, argue that car trips are only completed when the users expect that they get a net benefit from the trip; therefore, more car traffic demonstrates that the people are responding rational to there increased opportunities. Traffic growth is furthermore driven by exogenous factors.

In the UK and US, the authorities handle the topic quite fundamentally and released well-known studies, such as the SACTRA Report (1994). In the US of the late nineties, a number of studies were completed using pooled time-series/cross-sectional analysis on county or higher aggregated levels with the aim to distinguish between the influence of supply improvement and exogenous factors on transport demand. In most of these studies, the changes in trip making caused by supply improvement are called *induced travel* or *induced traffic*. Though there is some confusion about the word "induced" (*induziert*) in the German-speaking world, the English meaning is well defined in the literature. As Lee, *et al*. (1999) explain, "Induced is a term implying that a particular condition is indirectly caused by another condition". Similarly, Pickrell (2001) defines *induced* as, "...brought about through an indirect influence".

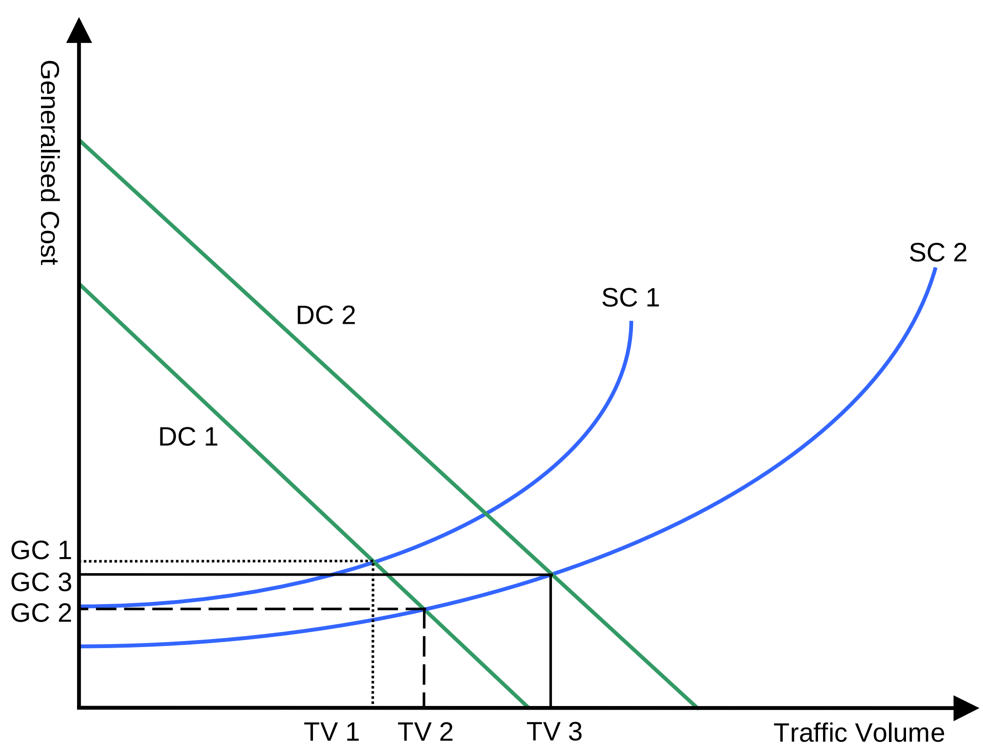
The remainder of this paper is organised as follows. Chapter 2 reviews the relationship between supply and demand, as well as the response of users to improvements in transport infrastructure. Next, Chapter 3 describes and analyses explanatory variables and econometric models for this topic. Finally, Chapter 4 critically examines existing studies with a view towards future work.

2. Induced travel and demand

Induced traffic means, that a road expansion, leading to lower generalised costs, results a higher demand afterward. The concept can be explained with a supply and demand diagram. Figure 1 shows that a supply curve, *SC 1*, increases exponentially due to increasing time costs on congested roads. These time costs form the majority of the generalised costs; other components (e.g., vehicle operation costs) do not increase as much with congestion levels. The corresponding demand function, *DC 1*, is the distribution sorted by the user willingness to pay.

An expansion of the capacity of a road leads to lower generalised cost, mainly because of minor time costs. Therefore, when the supply curve shifts downward, it meets the demand curve at a new point. This yields to a higher number of traffic volume (*TV 2*) and to lower generalized costs (*GC 2*). When demand increases, for example due to higher personal income, then the demand curve shifts from *DC 1* to *DC 2*, and the new equilibrium point leads to higher traffic volume (*TV 3*), but also higher generalized costs (*GC 2*) for the users.

Figure 1 Supply and Demand Curves for Generalised Cost and Traffic Volume



The possible impacts of lower generalised costs, driven by an increase of road capacity measured in *lane miles* (*LM*), on travel behaviour are:

1) a time of day shift (does not increase *daily vehicle miles of travel*, *VMT*)
2) route shift (may increase VMT)
3) transport mode shift
4) change in destination choice (longer trips)

   and

5) newly generated trips.

Cervero and Hansen (2002) and Noland (2001) summarise all five as induced travel and the latter three more specifically as induced demand.

Furthermore, there is a distinction regarding the timeframe. In the short run (less than one year) most factors stay constant, with the exception of generalised costs. Users do not have enough time to rearrange their employment or residential location. In the case of long run effects (more than 1 year), people choose their location with regard to the improvement of the supply side.

The goal of examining the elasticity of supply (road investment) can be stated as, are lane-mile additions significantly explained by changes in demand (car traffic)? (Cervero and Hansen, 2002)

3. Methodology and required data for aggregate econometric models

In the studies (Hansen and Huang (1997); Noland and Cowart (2000); Fulton, *et al.* (2000); Noland (2001); Cervero and Hansen (2002)) presently under examination, the sole element of analysis is the state roads (all roads in the system above the local level). On these roads, the VMT can be estimated with traffic counts. In order to estimate the traffic volume on non-state roads, Hansen and Huang (1997) used gas sales figures on county and metropolitan levels, respectively. Though the results were not statistical significant, they suggest that higher traffic volume on state highways does not reduce VMT on local roads.

Regional aggregation is analysed in these studies at the county, metropolitan, and/or state level. It is important to disaggregate, to the extent possible, and to verify that all changes in the transport system are intrazonal. Otherwise, there can be spill-over effects. For the planned data model for Switzerland, this means to aggregate the spatial planning regions into 30 to 50 regions.

The main goal of the single-equation method is to estimate the dependent variable, Y_{it}, with the independent variable(s).

$$Y_{it} = c + \lambda^k X_{it}^k + u_{it}$$

A common problem arising in cross-sectional data is hetroscedasticity (unequal variances); for example, error terms of larger regions might have larger variances than those error terms associated with smaller regions. To handle hetroscedasticity, the equation is transformed into a natural logarithm. The transformation shortens the magnitude of variance by a factor of 5.

$$\ln(VMT_{it}) = c + \sum_k \lambda^k \ln(X_{it}^k) + u_{it}$$

c	intercept
VMT_{it}	daily vehicle miles of travel in region i in year t
λ^k	parameter for the kth independent variable
X_{it}^k	independent variable k of region i in year t
u_{it}	error term of region i in year t

Furthermore, the slope parameter, λ^k, in the log form measures the point elasticity of VMT regarding the explanatory variable, X^k, directly and is defined as:

$$\lambda^k = \frac{\partial \ln(VMT)}{\partial \ln(X^k)} = \frac{X^k}{VMT}\frac{\partial (VMT)}{\partial (X^k)}$$

In the *ordinary least squares* (*OLS*)-estimation, one of the assumptions is that the estimation has a constant intercept and slope. In pooled time series/cross-sectional data, this restriction must be loosened, and dummy variables are introduced to allow the intercept term to vary over time and cross-sectional units, respectively. The so-called *fixed effect approach* (Pindyck and Rubinfeld, 1998) captures the influence of factors unknown or unmeasured in the model. Hansen, *et al.* (1993; as cited in Cervero, 2002) use this technique for the first time in induced demand analysis.

$$\ln(VMT_{it}) = c + \alpha_i + \beta_t + \sum_k \lambda^k \ln(X_{it}^k) + u_{it}$$

α_i fixed effect for region i

β_t fixed effect in year t

Another problem we face particularly in time series is *autocorrelation*, occurring when the error terms are affected by trends. The most popular test for auto correlation is the *Durbin-Watson test*. To solve the autocorrelation problem, the generalized difference equation, or *Prais-Winsten equation* (*P-W*), is used (Gujarati, 1995; Pindyck and Rubinfeld, 1998).

$$\ln(VMT_{it}) - \rho \ln(VMT_{i(t-1)}) = \alpha_i(1-\rho) + \beta_t(1-\rho) + \sum_k \lambda^k \ln(X_{it}^k - \rho X_{i(t-1)}) + \varepsilon_{it}$$

ρ portion factor

A further question is whether it is possible to isolate the growth in VMT for each of the different road types. The *seemingly unrelated regressions* (*SURE*; Zellner, 1962) can be employed with more than one equation simultaneously and where the endogenous variables bear a close conceptual relationship to each other. If there is no correlation between the error terms, the result is equal to the OLS estimation.

Multicollinearity occurs when at least two independent variables are highly correlated with each other. Possible indications are (Pindyck and Rubinfeld, 1998):

- a good fit of the regression line with the date (*high* R^2) is found along with a low significance for the independent variables,
- pairs of the independent variables are highly correlated,

and

- formal tests indicate multicollinearity, (c.f., Greene, 2000).

For data such as that employed here, two techniques can help to solve the multicollinearity problem. The first, combination of cross-sectional and time series data, is commonly used irrespective of multicollinearity. Transformation of variables is simply the *first difference form* (*FDF*), which reduces multicollinearity even if two independent variables are highly correlated, as this correlation does not indicate a high correlation in their differences.

$$\ln(VMT_{it}) - \ln(VMT_{i(t-l)}) = c + \alpha_i + \beta_t + \sum_k \lambda^k (\ln(X_{it}^k) - \ln(X_{i(t-l)}^k)) + \varepsilon_{it}$$

In economics, the influence of an independent variable on the dependent variable is hardly instantaneous. If on the right-hand side there is a series of lagged independent variable accounting for the time-adjustment process, the technique is called a *distributed-lag model* (*DL*). On the other hand, presence of a lagged dependent variable on the right-hand side is *autoregressive*, or *partial adjustment* (*PAM*), model. Distributed-lag models were frequently applied in this area, but Noland and Cowart (2000), and Noland (2001) used a partial adjustment model. Lags appear because of psychological, technological or institutional reasons (Gujarati, 1995).

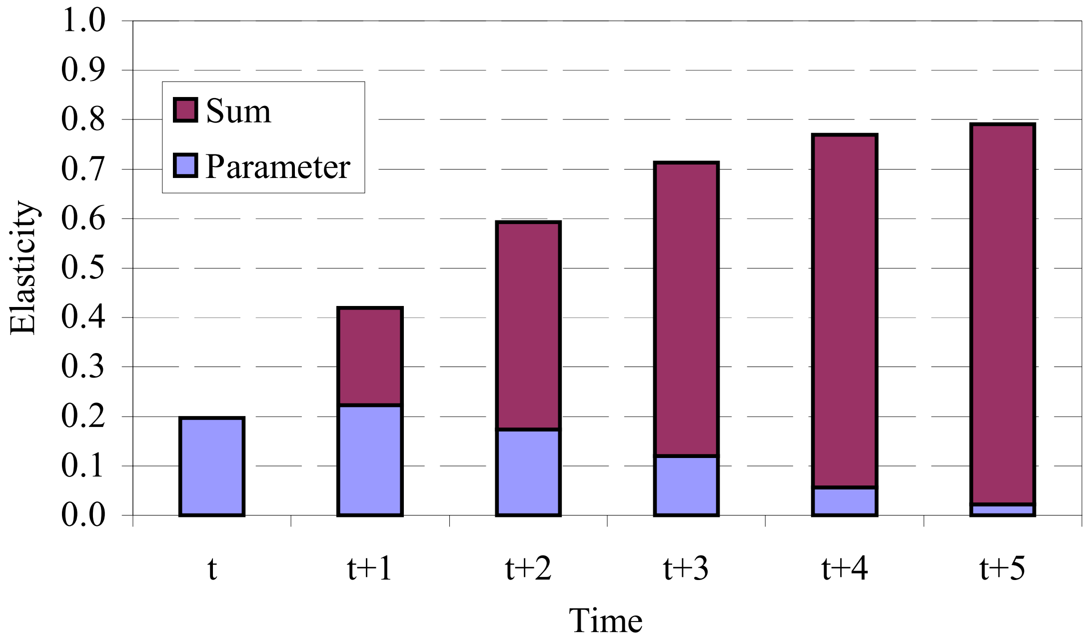
The *Koyck* approach (Gujarati, 1995) of the distributed-lag model uses a declining geometric distribution as the lag lengthens:

$$Y_{it} = c + \alpha_i + \beta_t + \lambda_0 X_{it} + \lambda_1 X_{i(t-1)} + \lambda_2 X_{i(t-2)} + \ldots + \lambda_k X_{i(t-k)} + u_{it}$$

The *polynomial distributed-lag* (*PDL*), or *Almon*, model (Almon, 1965) provides a more flexible method of using different possible forms in the lag structure. Cervero and Hansen (2002) found that a quadric polynomial form best fit the data, resulting in the elasticity values shown in Figure 2.

$$Y_{it} = c + \alpha_i + \beta_t + \sum_{l=0}^{k} (a_0 + a_1 l + a_2 l^2) X_{t-l} + u_{it}$$

Figure 2 Results of Polynomial Distributed-Lag Model



Using data from Cervero and Hansen (2002, p. 485)

In *simultaneous-equation estimation,* several equations are taken into account, and the behaviour of the variables is jointly determined. The model has more than one endogenous variable (determined within the system of equations), and, in the different equations, one or more of the explanatory variables (right-hand side) will be endogenous. This means it will be correlated with the error term.

To handle the described relationship, the *two-stage least squares method* (*2SLS*; Theil, 1953; Basmann, 1957) is employed. First, an *instrumental variable* (*IV*), a linear combination of the predetermined (exogenous or lagged endogenous) model variables, is estimated. Second, the estimated IV is used in the equation on the right-hand side.

However, the error terms between equations could be correlated. Therefore, the *three-stage least squares technique* (*3SLS*) is used (Pindyck and Rubinfeld, 1998). In this case, a generalised least-squares estimation is applied to a model of equations, first estimated by 2SLS.

In the *Granger causality test* (Granger, 1969), the relationship between cause and effect can be tested by means of quantitative methods; in other words, the cause for any given effect must precede that effect. A variable, X, is for a variable, Y, causal, if the forecast for Y is influenced by and improved with lagged values of X. Furthermore, the forecast of X may not be influenced by lagged values of Y. If these assumptions hold, it is a one-way causality.

Four cases are distinguished:

1) unidirectional causality LM ==> VMT, $\sum \alpha_i \neq 0$ and $\sum \delta_j = 0$

2) unidirectional causality VMT ==> LM, $\sum \alpha_i = 0$ and $\sum \delta_j \neq 0$

3) bilateral causality VMT <==> LM, $\sum \alpha_i \neq 0$ and $\sum \delta_j \neq 0$

4) independence $\sum \alpha_i = 0$ and $\sum \delta_j = 0$

In Table 1, the results for the Granger causality test are listed. Cervero and Hansen (2002) find a significant causality in both directions, with a higher significance in the direction demand (VMT) follows supply (LM). Fulton, *et al.* (2000) find a significant causality from supply to demand, but not in the other direction.

Table 1 Results of the Granger Causality Test

	Fulton, *et al.* (2000)	Cervero and Hansen (2002)
unidirectional causality LM ==> VMT	significant	significant
unidirectional causality VMT ==> LM	not significant	significant

Sources: Fulton, *et al*. (2000) p. 11; Cervero and Hansen (2002) p. 487;

Table 2 (on the following page) gives an overview of the applied data sets in the different studies as well as the variables used in the significant econometric models. Table 3 shows the best fitting results for the various models. Where possible (when disaggregated results were reported), the range is stated. The outcomes suggest short and long run elasticity of demand due to supply changes of 0.20 – 0.30 and 0.60 – 0.80, respectively. In terms of supply due to demand changes, the results suggest short and long run elasticity of 0.15 – 0.33 and 0.73, respectively.

Table 2 Overview key variables and data model specifications

Studies	Time period	Regional level	Numbers of regions	Supply var.	Demand var.	Population	Urban land area	Pop. density	Income p.c.	Employment	Fuel price	Fixed effect region	Fixed effect year
Hansen and Huang (1997)	1973 - 1990	CA Urban counties	30	LM	VMT	●			●		●	●	
	1973 - 1990	CA Metropolitan. Areas	14	LM	VMT	●			●		●	●	
Fulton, et al. (2000)	1970 - 1996	County level (MD,VA, NC, DC)	220	LM	VMT	●			●	●		●	●
Noland and Cowart (2000)	1982 - 1996	US Metropolitan area	70	LM/p.c.	VMT/ p.c.		●	●	●		●	●	
Noland (2001)	1984 - 1996	US States	50	LM/p.c.	VMT	●			●		●	●	
Cervero and Hansen (2002)	1976 - 1997	Cal. Urban counties	34	LM	VMT	●			●	●	●	●	●

Sources: Hansen and Huang (1997); Noland and Cowart (2000); Fulton, *et al*. (2000); Noland (2001); Cervero and Hansen (2002);

Table 3 Parameters for Elasticity of Demand

Studies	Method	Demand elasticity	Supply elast.	Population	Pop. density	Income p.c.	Employment	Fuel price	CO MAX, lag	Gov. Party 1=Dem, lag	Temp. diff.	White pop.
Hansen and Huang (1997)	P-W[1]	0.37		0.41		0.10						
	P-W[2]	0.53		0.47		0.16						
	PDL[1]	0.62 (l)[5]		0.46		0.05						
	PDL[2]	0.94 (l)		0.69		0.21						
Fulton, et al. (2000)	OLS	0.33 – 0.59		0.50 – 0.66		0.03 – 0.20						
	FDF	0.15 – 0.61		0.07 – 0.38		0.02 – 0.26						
	IV2	0.51		0.03		0.00						
	IV3	0.46		0.03		-0.03						
Noland and Cowart (2000)	OLS	0.65			-0.17	0.37		-0.05				
	IV	0.76			-0.16	0.34		-0.05				
	PAM	0.28 (s)[4]			-0.08 (s)	0.09 (s)		-0.03 (s)				
		0.90 (l)			-0.25 (l)	0.29 (l)		-0.08 (l)				
Noland (2001)	PAM	0.12 (s)		0.30 (s)		0.38 (s)		-0.05 (s)				
		0.37 (l)										
	FDF[3]	0.58 – 0.78		0.20 – 0.92		0.20 – 1.12		0.00 – -0.04				
Cervero and Hansen (2002)	3SLS	0.59		0.69		0.29	-0.08	-0.18				
			0.33	0.52			-0.32		0.06	0.01	0.50	0.41
	PDL	0.20 (s)		0.7		0.23	-0.11					
		0.79 (l)										
			0.15(s)	0.61			-0.37		0.04	0.05		
			0.73(l)								0.71	

Sources: Hansen and Huang (1997); Noland and Cowart (2000); Fulton, *et al*. (2000); Noland (2001); Cervero and Hansen (2002);

[1] county level [2] metropolitan level [3] values of urban collector roads omitted [4] short run [5] long run

4. Conclusion and outlook

The results in Table 3 show consistent parameters for elasticity of demand in different econometric models. Still, a weak point is the supply variable lane-mile (LM). However, adding two lanes in an urban area with highly congested roads will result another improvement for the car users in this area than it will in a rural area with low traffic. If one measures the lane mile elasticity of demand, it is (Lee, *et al*., 1999; Pickrell, 2002; Cervero 2002):

$$\lambda_{LM} = \left(\frac{\partial(VMT)}{\partial(v)}\frac{v}{VMT}\right)\left(\frac{\partial(v)}{\partial(LM)}\frac{LM}{v}\right)$$

where: VMT is demand
v is speed of travel
LM is supply

The first term in parentheses is the travel time elasticity of demand, which has a positive sign, and the second term is the travel speed elasticity of highway service, also with a positive sign. An important task in further work will be to find a better variable to measure the changes on the supply side.

A number of influences captured by the fixed effect variables in the studies would be intriguing to analyse further. These are the female participation in the workforce, the car ownership ratio, and supply and demand for public transport, all at the municipality or district level. Public transport has a much higher standard in Switzerland than in the US. Thus, in estimating the change in car traffic demand in response to road supply changes, the quality of public transport and the responding demand are also important.

At the IVT, a huge effort is presently put into the production of consistent time-series data for these variables. Furthermore, availability of gasoline price and sales numbers at the district level would improve the elasticity estimations. No appropriate source for this data has yet been found.

Analysis of previous work suggests the following key points regarding the Swiss data model. First, time-series data for the period from 1970 to 2000 should be collected. Additionally, the regional aggregation should be divided into four levels: spatial planning regions (138), aggregated spatial planning regions (approximately 30-50), and metropolitan regions (10). With the additional explanatory variables, it should be possible to build a more accurate model for Switzerland than those existing for the US. With this model, different hypotheses about traffic supply and demand over time can be tested.

References

Almon, S. (1965) The distributed lag between capital appropriations and expenditures, *Econometrica*, **33**, 178-196.

Basmann, R. (1957) A generalised classical method of linear estimation of coefficients in a structural equation, *Econometrica*, **25**, 77-83.

Cervero, R. (2002) Induced demand: An urban and metropolitan perspective, *Working Together to Address Induced Demand*, ENO Transport Foundation, Washington D.C.

Cervero, R. and M. Hansen (2002) Induced travel demand and induced road investment, *Journal of Transport Economics and Policy*, **36** (3) 469-490.

Fulton, L., D. Meszler, R. Noland and J. Thomas (2000) A statistical analysis of induced travel effects in the U.S. Mid-Atlantic region, *Journal of Transportation and Statistics,* **3** (1) 1-14.

Granger, C. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 424-438.

Greene, W. (2000) *Econometric Analysis*, 4th edition, Macmillan, New York.

Gujarati, D. (1995) *Basic Econometrics*, 3rd edition, McGraw-Hill, New York.

Hansen, M, D. Gillen, A. Dobbins, Y. Huang, and M. Puvathingal (1993) The air quality impacts of urban highway capacity expansion: Traffic generation and land use change, Berkeley: University of California, Institute of Transportation Studies.

Hansen, M and Y. Huang (1997) Road supply and traffic in California urban areas, *Transportation Research - A,* **31** (3) 205-218.

Lee, D., L. Klein, and G. Camus (1999) Induced traffic and induced demand, *Transportation Research Record*, Vol. **1659**, 68-75.

Noland, R. and W. Cowart (2000) Analysis of metropolitan highway capacity and the growth in vehicle miles of travel, Presented at the 79th Annual Meeting of the Transportation Research Board, Washington D.C.

Noland, R. (2001) Relationships between highway capacity and induced vehicle travel, *Transportation Research A*, **35** (1) 47-72.

Noland, R. and L. Lem (2002) A review of the evidence for induced travel and changes in transportation and environmental policy in the US and the UK, *Transportation Research D,* **7** (1) 1-26.

Pickrell, D. (2002) Induced demand: Definition, measurement, and significance, *Working together to address induced demand*, ENO Transport Foundation, Washington D.C.

Pindyck, R. and D. Rubinfeld (1998) *Econometric Models and Economic Forecasts,* 4th Edition, McGraw-Hill, New York.

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

Theil, H. (1953) Repeated least-squares applied to complete equation systems, The Hague: The Central Planning Bureau.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, **57,** 348-368.