Applying Conceptually Different Models to One Behavioural System: Choices by Car Tourists

John H.E. Taplin
University of Western Australia

This paper is concerned with the implications of models with entirely different bases 'explaining' the same revealed behaviour. There are no general answers but specific studies of car tourism in part of Western Australia have been used to elucidate contrasts, parallels and complementarities. The models used to represent holiday destination choice were standard linear programming, microsimulation, nested multinomial discrete-choice analysis and gravity-assignment traffic modelling. The latter two, the analytical-statistical models, complement each other and also provide inputs for the synthetic models. On close examination, the simulation and discrete-choice models are found to provide close parallels in the ways they explain choice. The capacity of LP and simulation to give identical levels of explanation follows, to at least some degree, from the fact that goals and constraints work similarly in both, even though the LP is aggregate and simulation disaggregate. The major contrasts are the use of an imaginary collective goal for the LP, as compared with individual choice in the simulation, and the long planning horizon implied by the LP compared with the myopically short one of the simulation.

1. INTRODUCTION

Although the title of this paper is generic, the topic is approached through a set of applications to a specific case. Apparent goodness of fit is not an issue in this case; it is not difficult to make different synthetic models approximate to observed behaviour when there are almost unlimited degrees of freedom.

It is relatively common for essentially the same situation to be represented by two different models or for different models to overlap and be used to explain the same thing. A familiar example is neo-classical and discrete choice demand modelling. Recent examples of studies which compared forecasting models are by Nijkamp et al. [1996], comparing logit and neural network alternatives for modelling inter-urban travel, and by Smith and Demetsky [1997], comparing ARIMA, non-parametric regression and neural network for forecasting traffic flow.

A model is generally deemed satisfactory if it can be validated, often with little regard to the realism of the assumptions. This was noted by the famous economist Milton Friedman [1953], who pointed out that the worth of various general and partial microeconomic models is judged not by their assumptions but by how well they predict. This is a non-trivial distinction when one remembers such curiosities as identical individuals with identical tastes lurking in microeconomic models. It is from such beginnings that important demand relationships, symmetry and homogeneity are derived, but that does not weaken them at all. Symmetry has gained law-like status not because of rigorous derivation but because of empirical validation in a range of situations [Brown and Deaton, 1972; Barten, 1977; Blundell, 1988]. The validation is so strong that symmetry is expected to hold not only in neo-classical demand systems but also in systems based on discrete choice analysis. Should symmetry be lost during aggregation over individuals then it is reasonable to reimpose it [Taplin, Hensher and Smith, 1999].

If models stand on the basis of empirical validation rather than their analytical or conceptual basis then there is no fundamental objection to equally validated but conceptually different models of the same situation, even though they complicate one's understanding of both models and validation. The primary examples in this paper are a linear program and a microsimulation which have been used to reproduce, almost exactly, the survey results for a set of tourist choices. The capacity of an LP to positively represent or 'simulate' the collective effect of processes which are not centrally managed has been demonstrated a number of times. Knight and Taplin [1971] generated an agricultural supply function by modelling, through a range of prices, each farm in an industry survey sample. Recently, Murphy and Mudrageda [1998] have used LP to generate energy supply and demand curves.

Two statistical models are also presented because these complementary approaches contribute added insights into tourist behaviour. In all four cases, the procedure has been to obtain a 'fit' to the survey data. At a simplistic level, one can argue that, to the extent that the synthesised results are comparable to trip patterns revealed by survey data, they offer insights into tourist decisions.

The statistical analyses of observed outcomes provide estimates of general tendencies or preferences running through a population of travellers. The resulting parameters provide central values around which appropriate distributions can be constructed for simulation.

Each of the two synthetic models carries with it different implications about the realities of tourist travel behaviour. The LP implies optimising over a whole trip and also implies that choices are heavily constrained.

The simulation involves a sequence of local decisions driven by an array of random distributions of characteristics, influences and attractiveness. The implied behaviour is a serendipitous - or mindless and short sighted - progression through a sequence of choices. Perhaps that is overstating it but, of all the models, a simulation seems to have the least theoretical underpinning, apart from the relevant probability theory. It is a productive way of exploring complex joint probabilities.

But an important question for this paper is what to make of it if models with entirely different bases can be made to 'explain' the same revealed behaviour.

1.1 The Applications

The paper deals with separate studies, by the author and graduate students, of car tourist choices. The LP and simulation models deal with return car trips to the six main holiday centres on the north-west coast of WA. A major incentive for winter holidays is to seek warm weather. Figure 1 indicates the main differences between the models.

Figure 1 Summary Comparison of Model Attributes

	Multi-Period Linear Program	Discrete-Event Simulation	Discrete Choice	Gravity and Assignment
Broad type	Synthetic	Synthetic	Analytical	Analytical
Model level	Classes of tourists	Individual parties	Individual parties	Road link flows
Procedure	Standard linear program	Discrete event simulation	Multinomial logit (nested)	Genetic algorithm search & Newton
Goal	Maximise satisfactions	Daily preferences	Maximise utility	Maximise utility
Fitting method	Fit sum of weighted solutions to survey	Calibrate for best fit to survey	Max likelihood fit to choices	Max likelihood fit to road link flows
Constraints	Drive distances & holiday length	Available choices & holiday length	The assumed choice set	Road network and destinations
Behaviour	Enjoy planned holiday	Hedonistic daily	Utility maximisers	Attracted to dests.
Horizon	Whole trip	Today's decision	Long-run	Long-run

2. THE LINEAR PROGRAM

Standard LP was applied to individual classes of car tourists, each being defined by driving constraints and satisfaction weights. A class was treated as one initial unit to be progressively broken into fractions going to various destinations day by day. The daily driving constraint became an average for the class, so that some could drive further, with an upper limit determined by the longest point to point link in the specification. Details are given by Taplin and McGinley [1999].

On the morning of day one, an entire class of tourists starts from Perth and shares of these tourists are delivered to various destinations. Each share is then available on the next day either to stay or move to another destination. This goes on over the length of the holiday and an equality ensures that all tourists return to Perth. The LP finds the levels of activities - daily drives between centres and stays at centres - which maximise total satisfaction. The result is a day to day pattern of trips and recreation stops which is sensitive to the weight attached to the various types of satisfaction in the objective function. Each model specification gave a distribution of tourist trips and stopovers for a particular set of weights.

An implication is that tourists plan the whole day by day sequence of travel and stopovers before starting the trip. For a week long trip, this is probably realistic for most travellers. It does not imply any limitation on impulsive choice of activities at destinations.

One group of benefits comprised resort activities: swimming, fishing, boating, sightseeing and eating at restaurants. These benefits only accrue when there is a day's stopover, all of the benefits associated with the particular centre being counted. No attempt has been made to differentiate between travellers who do and do not participate in the various activities.

The second group of benefits comprised warmth - the lower the latitude the greater the benefit - and the satisfaction derived from visiting a variety of destinations. Half of the local warmth benefit has been attributed on arrival at a destination and a further full day benefit if there is a day's stay.

2.1 LP Results

Figure 2 shows, as an example, the day to day pattern for relatively high and low weights applied to the satisfaction of multiple destinations. The fitting procedure was to compute LP solutions for a variety of classes of tourists and then to form a weighted sum by applying contribution weights which minimised the sum of squared deviations of the aggregated LP outputs from the survey distribution. The composite result included 15 solutions covering six, seven and eight day trips (5, 6 and 7 nights).

The proportion of the sum of squared deviations from the mean share of nights over seven destinations and seven nights (one forty-ninth) 'explained' by the composite LP result is 96%.

The individual LP results underlying the synthesised distribution were affected substantially by variations in the average daily drive constraint, in the weight for desiring to visit more destinations and in the weight for seeking warmer destinations. The weights for individual objectives, swimming boating etc., were not varied and the global driving constraint had little effect.

Figure 2. Daily Destination Shares for Low and High Weights on the Satisfaction from Going to Multiple Destinations (All other weights = 1)

	Km from Perth	Latitude South	Destination Shares by Night of Trip [a]						
			Destns Wt = 1	Destinations Weight = 5					
			Each Night	**1**	**2**	**3**	**4**	**5**	**6**
Jurien Bay	228	30° 18'	0.52	0.12					0.28
Geraldton	423	28° 46'		0.88	0.12			0.28	0.65
Kalbarri	590	27° 43'	0.48		0.15				
Denham/Monkey Mia	858	25° 48'							0.07
Carnarvon	904	24° 53'			0.73	0.21		0.67	
Coral Bay	1133	23° 08'				0.12	0.67	0.05	
Exmouth	1274	21° 56'				0.67	0.33		

[a] Day/night 7 is not shown because it simply involves the return to Perth.

3. THE SIMULATION

The simulation choices are driven primarily by distributed tourist desires or goals interacting with destination characteristics:

Tourist Goals Each individual party is randomly assigned a score for the importance attached to water based activities, sightseeing, eating at restaurants, satisfaction from visiting a variety of destinations and satisfaction from going to warmer latitudes in winter.

Specific Destination Benefits Each destination has been assessed with respect to each of three benefits: water based activity, sightseeing and eating at restaurants. The fourth benefit is warmth, which is important for winter holidays to warmer latitudes. The lower the latitude the greater the benefit. It is the interaction between the destination attributes and the corresponding tourist attributes that drives the choices.

3.1 Utility Function for Destination Choice

The expression for utility is a sum of benefits of going to a destination divided by a sum of disutilities incurred in going there and a random variate. Thus, the utility for party i of choosing destination j,

$$U_{ij} = \frac{D_{js}x_{is}+D_{jw}x_{iw}+D_{jr}x_{ir}+D_{jh}x_{ih}C_h}{C_d(d_{ij-1}+C_{Pd}d_{ij-1P})+C_p y_{ipj}x_{im}+N(1,0.1)}$$

D_{js} - rating of destination j for sightseeing
x_{is} - desire of members of party i to see sights
D_{jw} - rating of j for water based activities
x_{iw} - desire to participate in water based activities
D_{jr} - rating of destination j for restaurants
x_{ir} - desire to eat at restaurants
D_{jh} - latitude of destination j converted to a scale from 0 for Perth to 1.0 for Exmouth
x_{ih} - desire to seek a hotter climate
C_h - weight on desire for a hotter climate; for testing responsiveness; normally set to 1
d_{ij-1} - distance of j from the previous destination
d_{ij-1P} - increased distance of j from Perth
C_{Pd} - proportion of added distance from Perth that is perceived as a deterrent
C_d - disutility per kilometre of driving
y_{ipj} - 1 if j visited previously, 0 otherwise
x_{im} - desire to visit multiple destinations
C_p - discount for previously visited destination
$N(1,0.1)$ - random normal variate with mean of 1 and standard deviation of 0.1.

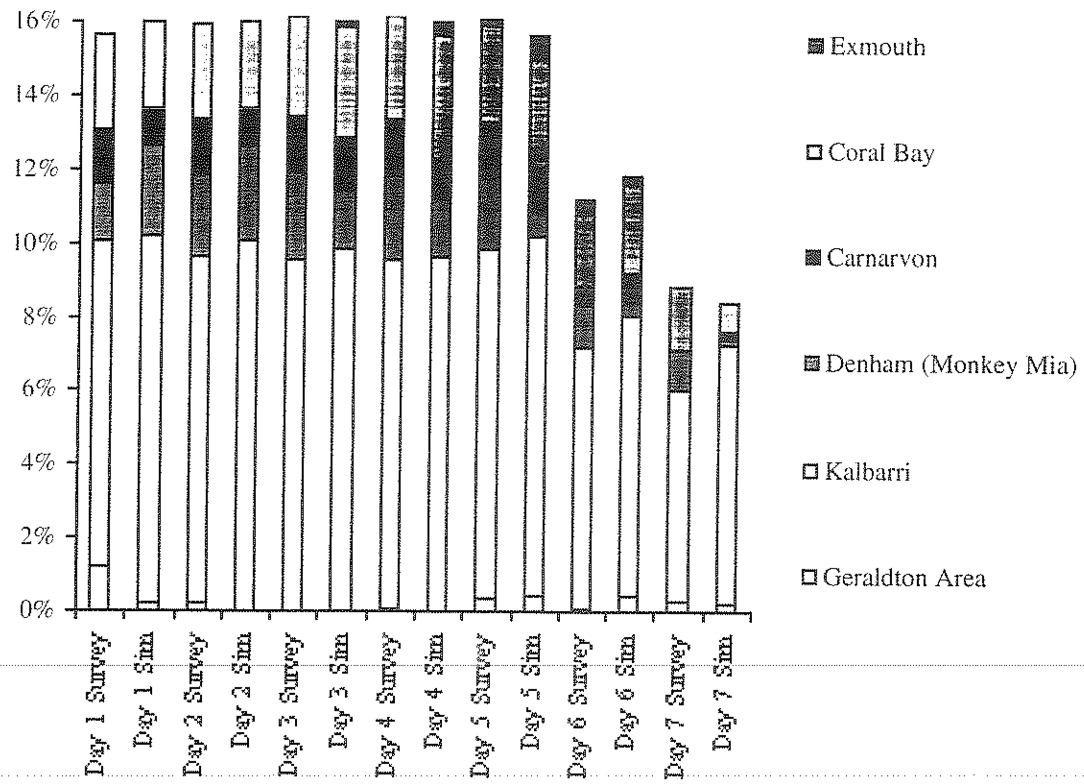
Modified and simpler decision functions are used for destination choice on Day 1 and for the decision to stay another day or not.

3.2 Simulation Results

Figure 3 presents a comparison of actual survey shares for the whole seven nights with the shares resulting from 400 simulated trips. There is some under-representation of Carnarvon and sometimes over-representation of Kalbarri in the simulation based distribution.

A pseudo-R^2 of 0.96 was obtained for squared deviations from the global mean share of nights over six destinations and seven nights 'explained' by the composite simulation result.

Figure 3. Comparison of Survey and Results for 400 Simulated Trips: Destination Shares of Total Trip Nights



4. THE DISCRETE CHOICE MODEL

The discrete choice model uses survey data for the whole of Western Australia and deals with the eight top non-urban tourist destinations, relating destination shares to the attributes of these places. A high perceived quality of a national park in an area may give a high probability of a visit and a lower probability of visiting other areas. This model of tourist choice is based on the availability of activities, accessibility and characteristics of the travelling party. The modelled groups were those in the 1995 and 1996 Western Australian Travel Survey (WATS) which satisfied the following criteria:

- Stayed at least once at one of the 8 destinations
- Used car or 4-wheel drive
- The main purpose of travel was pleasure/holiday
- The group was not part of an organised tour.

Approximately one hundred and fifty group visits to each of the eight destinations were randomly selected for analysis, from the groups meeting the criteria, giving a sample of almost 1,200 groups.

The rating of each activity attribute for each destination was taken to be the percentage of tourists undertaking the activity there. Three attributes, fishing, beaches and rivers, and boat harbours and estuaries, were taken to be generic and to have the same impact for all towns. Other attributes were treated as alternative specific.

4.1 Discrete Choice Results

Generic results and specific results for the four northern destinations are shown in Figure 4. On average, a prediction of the town visited would be correct in 49% of cases. The derived elasticities of destination choice with respect to various distance classes are shown in the last column of Figure 4.

The disincentive effect of distance on destination choice is greatest in the case of distance from home. One-way distances in the range of 300-600 kilometres from the last destination are also important influences on choice. The more attractive a destination, however, the less the relative impact of distance.

Apart from distance, the three generic variables were significant influences on choice. Various, generally plausible, destination specific influences were also estimated to be significant.

Figure 4. Significant Coefficients Estimated with Choice Model for the Eight Top WA Destinations: Only Northern Destinations Shown

	Attribute	Coefficient	't' ratio	Elasticity of Destination Choice w.r.t. Distance
Generic Variables	Dist. from last dest. (first 300 km)	-0.001709	-2.05	-0.19 to -0.23
	Dist. from last dest. (300-600 km)	-0.008026	-10.26	-0.39 to -0.52
	Dist. from last dest. (>600 km)	-0.004276	-6.56	-0.06 to -0.20
	Distance from home per night	-0.004888	-7.51	-0.55 to -0.69
	Fishing	3.51	2.60	
	Boat Harbour/Estuary	12.14	5.73	
	Beach or River	2.94	2.24	
Specific - Kalbarri	Museum or historical site	14.43	2.78	
	National Park	5.93	6.57	
	Wildlife Park/Fish Hatchery/Zoo	4.34	2.69	
	River activity	12.11	2.53	
Specific - Denham-MM	Dolphins	4.99	7.11	
	Museum or historical site	19.12	3.26	
	National Park	8.55	4.33	
Specific - Exmouth	Museum or historical site	12.87	2.44	
	National Park	7.31	4.70	
	Boating	7.54	2.87	
Specific - Broome	Museum or historical site	11.77	2.83	
	Locally organised tours	13.73	2.08	
Constant (Broome = 1)	Kalbarri	-5.00	-6.53	
	Denham/Monkey Mia	-4.31	-5.77	
	Exmouth	-2.79	-4.47	
	Number of observations	1183		
	ρ^2	0.518		

5. THE GRAVITY-ASSIGNMENT MODELS

The capacity of tourist destinations to attract visits and the propensity to make round trips to remote destinations were approached through models based on populations, travel times, traffic on road links and identification of prime tourist destinations. The primary data for these studies were road link flows in a limited rural network. Each model estimated a population multiplier for recognised tourist destinations.

For the North-West, genetic algorithm was used to simultaneously estimate a gravity model of trip generation, incorporating a tourist attraction population multiplier, and a logit route assignment model [Taplin and Qiu, 1997]. For this study, 29 internal origin/destination zones were defined and 7 external zones.

For the Southern Wheatbelt, maximum likelihood estimation of the combined gravity and logit route choice model started with genetic algorithm, leading to an approximate solution, from which the quadratic hill-climbing variant of the Newton then found the optimum, driving the first derivatives to zero. Using genetic algorithm as the first stage in the solution procedure gives a high degree of assurance that the result is not a sub-optimum [Han, 1998]. There were 22 zones within this region and 17 external.

Figure 5. Goodness of Fit and Implied and Estimated Elasticities

	MODEL FIT	ELASTICITY			Tourism Multiplier
		Distance	Destination Choice w.r.t dist.		
			to next dest.	from home	
Multi-Period LP	Pseudo-R^2 = 0.96				
Discrete-Event Simulation	Pseudo-R^2 = 0.96	-0.54 (implied)			
Discrete Choice	$\rho^2 = 0.53$		-0.19 to -0.52	-0.55 to -0.69	
Gravity & Assignment					
Wheatbelt South	$R^2 = 0.95$	-1.65			1.56
North-West	$R^2 = 0.96$	-2.18 to -3.22			3.83

6. IMPLICATIONS OF MODELLING ALTERNATIVES

Goodness of fit and global parameters are shown in Figure 5. The estimates are complementary rather than alternatives.

Although the four methods fall naturally into synthetic and analytical/statistical groups, there is a strong affinity between the simulation and discrete choice models. Both depend on the interaction between party desires and destination attributes. This reflects the fact that they work at the level of the individual party. One model applies attributes to parties according to predetermined probabilities, depending on calibration to achieve trip patterns which approximate to observed data, and the other calibrates the interaction parameters to fit the individual records. Key similarities and differences are:

	Goals	**Constraints**
LP	Maximise global satisfaction	Drive distances & holiday length
Simul'n	Max daily utility: interaction of desires and dest. attributes	Available daily choices, drive distances & holiday length
Choice	Max utility sum: interaction of desires and dest. attributes	Choice set (destination attractions) & drive distances
Gravity -Assign	Maximise utility	Road network and destinations

In the area studied, little was known about the relative importance attached by car tourists to the satisfaction of various goals and types of enjoyment. Thus, the allocation and manipulation of relative satisfaction weights has been arbitrary but the changes in both simulation and LP results in response to these manipulations give some indication of what should be probed in future survey work.

6.1 LP and Simulation Alternatives

In this application, the LP is essentially simulating group behaviour. Whereas the simulation takes each party individually, assigns characteristics and preferences randomly and then uses these to determine their choices, the LP determines the best sequence of choices for a class of tourists. Because it merely imposes an average daily driving distance, some members of the class drive short distances and others drive long distances.

At the micro level, the simulation retains the identity of each party throughout, so that each individual itinerary can be studied at the end of the simulation, but the LP only occasionally parallels this. It does break a class of tourists into sub-groups and one can sometimes track itineraries through the day to day LP results. However, individuality is lost when, for example, two sub-groups go to different destinations on day two, then on day three converge on another destination, from which two sub-groups diverge again on day four. The day four sub-groups cannot be matched to the day two sub-groups. In principle, integer programming could be applied to individual parties but that was not attempted.

The simulation was fitted to the aggregate survey data by arbitrarily varying parameters until the destination outcomes gave a close approximation to the survey outcomes. However, the solution is not unique and it is probable that other parameter settings could do as well. At the level of destination aggregates, the solution is over-determined because the number of parameters which can be varied exceeds the number of observations. A more appropriate procedure would be to fit the individual party itineraries to the recorded survey party itineraries.

In the case of the LP, the strength of the multiple destinations and warmth goals were varied, and also the average daily driving constraint and the maximum daily drive (by specifying which point to point drives were available). Each resulting solution represented a class of tourist. The final solution was a weighted sum of fifteen of these class solutions. Again, an over-determined solution, particularly if some of the other goal weights were to be varied.

7. CONCLUSIONS

The synthetic and analytical models fairly clearly complement each other. A synthesised model seems to be a natural way of approaching choice processes from the opposite side to the ex post analysis of choices after they have occurred in real life. The latter seeks to identify factors which have contributed to choice whereas the synthetic model seeks to mimic the structure of choices as they may have occurred. There is no tension between the two approaches. In due course, a properly designed stated choice or preference survey, along the lines of Louviere and Hensher [1983] and Hensher, Louviere and Swait [1999], will be used to probe decision factors more fully and thus go a considerable way to filling the gap in our understanding of decision processes that the simulation and LP models have sought to fill.

Between the analytical models themselves there is a contrast but no conflict. The nested logit analysis is based on a rich data set recording individual trips, with destinations, activities, transport and accommodation. In contrast, the gravity-assignment model is based on minimal data, traffic counts and populations. From these, fairly reliable estimates have been made of the attractiveness of tourist destinations, in relation to normal centres, as well as estimated elasticities of demand for travel distance. These augment what has been estimated from the survey data.

However, there is a tension or contrast between the two synthetic models, simulation and LP. These are truly alternative representations of exactly the same tourist choice behaviour. Despite the differences in formal method, much of what the two models do is similar. A major difference that remains is the level at which decisions are made and goals are satisfied. The simulated decisions are made by the individual party at the local level without regard for subsequent decisions, except to ensure movement back towards Perth on the second last day. In contrast, the LP 'decisions' are global and the weighted sum of goals is maximised collectively. This may not be as inappropriate a representation of individual goal seeking as appears at first. It is a class of tourists that is optimised collectively and the implied forward planning may be as consistent with behaviour on a week-long vacation as the day-to-day 'what do we do now' behaviour implied by the simulation.

In brief summary, the analytical-statistical models not only complement each other but also complement and provide inputs for the synthetic models. Furthermore, the simulation and discrete-choice models provide extremely close parallels in the ways they explain choice. The capacity of LP and simulation to give identical levels of explanation is not surprising. Goals and constraints work similarly in both, even though the LP is aggregate and simulation disaggregate. The only major contrasts are the use of an imaginary collective goal for the LP, as compared with individual choice in the simulation, and the long planning horizon implied by the LP compared with the myopically short one of the simulation.

8. REFERENCES

Barten, A.P., The systems of consumer demand functions approach: a review, *Econometrica*, 45, 23-51, 1977

Blundell, R., Consumer behaviour theory and empirical evidence - a survey, *Economic Journal*, 98, 16-65, 1988

Brown, A., and A. Deaton, Surveys in applied economics: models of consumer behaviour, *Economic Journal* , 82, 1145-1236, 1972

Friedman, M., The methodology of positive economics, *Essays in Positive Economics*, Chicago UP, 1953

Han, R.L., Calibration of parameters for a combined gravity and traffic assignment model, *Proc. of the Optimization Mini-conference (V)*, Perth, 1998

Hensher, D., Louviere, J., and J. Swait, Combining sources of preference data, *Journal of Econometrics*, 89, 197-221, 1999

Knight, R.C., and J.H.E. Taplin, Supply response in the peanut industry: a linear programming approach, *Quarterly Review of Agricultural Economics*, 24, 112-122, 1971

Louviere, J.J., and D.A.Hensher, Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event, *Journal of Consumer Research*, 10(3), 348-361, 1983

Murphy, F.H., and M.V.Mudrageda, A decomposition approach for a class of economic equilibrium models, *Operations Research*, 46(3), 323-338, 1998

Nijkamp, P., Reggiani, A., and T.Tritapepe, Modelling inter-urban transport flows in Italy: a comparison between neural network analysis and logit analysis, *Transportation Research Part C*, 4(6), 323-338, 1996

Smith, B.L., and M.J. Demetsky, Traffic flow forecasting: comparison of modeling approaches, *Journal of Transportation Engineering*, 123(4), 261-266, 1997

Taplin, J.H.E., and C. McGinley, A linear program to model daily car touring choices, *Annals of Tourism Research*, 27(2), 1999

Taplin, J.H.E., Hensher, D.A., and B.Smith, Preserving the symmetry of estimated commuter travel elasticities, *Transportation Research Part B*, 33(3), 215-232, 1999

Taplin, J.H.E., and M. Qiu, Car attraction and route choice in Australia, *Annals of Tourism Research*, 24, 624-637, 1997