Forecasting Travel on Congested Urban Transportation Networks: Review and Prospects

David E. Boyce* Huw C. W. L. Williams†

*Department of Civil and Environmental Engineering, Northwestern University
2149 Grey Avenue, Evanston, IL 60201 USA
dboyce@uic.edu

†Department of City and Regional Planning, Cardiff University,
Cardiff CF10 3WA, UK
WilliamsH@Cardiff.ac.uk

1 Introduction

Fifty years ago, Beckmann, McGuire and Winsten (1956) completed their seminal formulation and analysis of an integrated model representing origin-destination, route and link flows on a network, as a function of flow-dependent link costs. Their treatment of this fundamental problem, variously known as the traffic assignment problem with variable demand, the multi-commodity network flow problem and the combined model of trip distribution and traffic assignment, sparked a vast literature numbering well over 1,000 references (Patriksson, 1994).

At the same time transportation engineers and planners in Detroit, and then in Chicago, were grappling with a way to solve computationally essentially the same problem with primitive electromechanical accounting machines and very early computers, in order to fashion future road and transit system plans. Unaware of the significance of the above formulation, they proposed a method, now known as the four-stage or sequential procedure, which amounts to a simple heuristic for solving the integrated model. By the time the significance of the model of Beckmann et al was understood, and a solution algorithm devised, the sequential procedure became so widely accepted that it is now widely regarded as effectively axiomatic.

This paper examines the history of these developments in both research and practice, and seeks to understand, clarify and interpret why these events occurred as they did. Then, an effort is made to step back and examine whether and how progress in travel forecasting practice can best be advanced in the future. The paper consists of six parts: 1. Origins of Travel Forecasting Models; 2. Four-Stage Paradigm and Its Development; 3. Emergence of Combined Models; 4. Current Status of Integrated Models; 5. Applications and Related Software Development; 6. Future Prospects.

2 The Origins of Travel Forecasting Models

Traffic congestion is a phenomenon that mankind finds wasteful and offensive. So it was in the early 1950s when mathematical economics, a new and rapidly developing field, sought to tackle practical problems. A team of young economists (Beckmann, McGuire and Winsten, 1956) took up the problem of congestion in a transportation network, and succeeded in devising a mathematical model of travel and route choices that contributed in a fundamental way to this new field. Assuming that travel between each pair of origins and destinations decreases with increasing cost, that used routes connecting each pair have minimal and equal travel costs, and that the travel cost on each link of the network is an increasing function of the total link flow, Beckmann devised an optimization problem whose solution simultaneously satisfies these three conditions. Although the solution properties of the problem were thoroughly analyzed, no solution method was devised, except for a simple procedure applied to a tiny network.

This work, completed in 1954, was published as *Part I, A Study of Highway Transportation*, only in 1956 and never appeared in academic journals. Evidently, the book was widely distributed, as it had three printings plus a Spanish edition. Even so, the book did not impact the urban transportation studies that began in those same years, perhaps because the mathematical treatment was not accessible to the engineers and planners who staffed those agencies. As a result, one of the most important innovations of this field was effectively lost for over ten years, by which time a more pragmatic travel forecasting paradigm had taken hold.

3 The Four-Stage Paradigm and Its Development

In place of this integrated model of travel and route choices proposed by Beckmann, a four-stage travel forecasting procedure evolved; see Ortúzar and Willumsen (2001) for a modern exposition of the procedure. Its origins are obscure, but Martin, Memmott and Bone (1961) thoroughly described the procedure in 1961, when they depicted it in a complex, multi-page diagram. At its heart lay four stages or steps: trip generation (G); trip distribution (D); modal split (MS) and traffic assignment (A). Each was depicted as a separate stage that received inputs from the former and provided outputs to its successor.

Following the acceptance of the four-stage paradigm, most researchers and professionals became engaged in the improvement of the models and methods described in the individual stages. Household-based category analysis replaced zone-based regression models (Wooton and Pick, 1967). Variously described utilities or generalised costs emerged from early studies of modal choice with models specified and estimated at the individual level. The incorporated generalised costs, specified as linear functions of objectively measured attributes with travel time suitably scaled to money units, served as an interface between policies, behavioural response and benefit evaluation. The numerical estimate of the *value of time* has proved to be one of the most important parameters in the whole of planning.

From an analytic viewpoint, the earliest distribution and modal split models, which involved apportioning trips between different locations and modes, adopted empirically derived functions – sometimes referred to as deterrence functions (for spatial interaction), and diversion curves (for modal shares) – and these were determined through *goodness-of-fit* criteria. By the late 1960s these began to be replaced in academic discourse and some applications by analytic functions, and share models of the multinomial logit form became widely adopted (Wilson, 1970; Manheim, 1979). These were conceptually appealing, analytically tractable and consistent with a number of theoretical constructs that were starting to be used for interpreting dispersion associated with trip making (Erlander and Stewart, 1990). These analytic expressions more readily allowed for an appreciation that their *sensitivity* parameter(s) served a four-way role in determining the dispersion in travel patterns, the frequency distribution of trip lengths with respect to generalised cost, the response to transport policies, and economic benefit measures.

A problem that exercised the earliest modellers was the ordering of the G, D, MS and A segments and how they should be *linked* together. There were, from the start, informal behavioural assumptions underpinning the four-stage approach in terms of a sequence of decisions that mapped onto the individual submodels. However, the correspondence between the Generation (G), Distribution (D), Modal Split (MS) and Assignment (A), with the frequency, location, mode and route choices of travel, respectively, remained tenuous.

Various alternative structures for the demand model were proposed reflecting, it was assumed, the conditionality of a sequence of decisions, the most popular being whether the distribution submodel preceded (G/D/MS/A), followed (G/MS/D/A) or was combined (G/D-MS/A) with the modal split submodel. The first two constructions involved the formation of "composite costs" that combined costs at what were referred to as "later stages of the models." As late as the mid

1970s no detailed theoretical basis for the entire model existed; for a given ordering, which was suggested *a priori* from behavioural assumptions, the form of the composite costs were regarded as extra degrees of freedom for achieving improved "goodness-of-fit."

The derivation of the nested logit model within discrete choice theory provided one resolution to these ambiguities (McFadden, 1973, 1978; Williams, 1977; Daly and Zachary, 1978). This development endowed the whole model with a behavioural rationale in which the analytical structure of the demand function reflected underlying utility functions, imposing two important restrictions on the overall model. Firstly, the composite costs that interfaced the different submodels needed to be formulated in a particular way; for logit-type models these were in the form of a "log sum" function, a form that had already been implemented with microdata by Ben-Akiva (1974). Secondly, the parameters that determined the sensitivity of travel choices to changes in times or costs, had to decrease as one progressed from route choice, through mode to locational and frequency selection in the G/D/MS/A structure. Only then do the *estimated* direct- and cross-elasticity parameters have the appropriate sign, requiring the demand for an alternative to fall when its cost rose or the cost of a substitute fell. The nested logit model thus provided a consistent way of combining the various constituent choices with differential cross substitution between alternatives, and made the ordering of associated logit share functions subject to empirical test. It is important to note that the specification of the demand model with empirically derived functions for locational and/or modal shares is not immune from this strict requirement for appropriate response properties derived from the calibrated model.

Williams and Senior (1977) reconfigured the four-stage procedure as a nested logit structure, experimented with different orderings of the distribution and modal split models, and showed that many practical models in the UK did not satisfy the necessary parameter inequalities implied by the chosen structure. The specific implication of this result was that such models could have produced counter-intuitive results. The more general implication was that calibration and the traditional notion of validation based on goodness-of-fit and held-back samples was not a sufficient test for the validity of such cross sectional models for policy testing. We return to both these points below in examining the current status of widely-applied nested models.

4 Emergence of Combined Models

A fundamental problem that has confronted theoreticians and practitioners since the earliest days was the design of solution procedures to determine equilibrium states of required accuracy, in which the costs (or times) of travel through the networks are consistent with the demands that created them. The practical difficulties and protracted time taken to grapple with this problem almost certainly resulted from the statement of the problem that Beckmann et al had produced years earlier being unknown to the professionals who proposed the four-stage procedure.

Through the late 1960s the problem of equilibrium was almost entirely seen to be confined to the assignment stage to be solved for given trip matrices. Even this restricted problem generated several ad hoc procedures before rigorous solution algorithms based on the Beckmann formulation emerged. In the morass of numerical detail involved in handling several large trip matrices and networks, the additional complexity of seeking self-consistency throughout the entire procedure tended to be seen as an unnecessary luxury or was simply ignored.

Where it was considered, the notion of feedback of generalised costs from the assignment to other stages of the model began to be discussed. The congested costs were simply recycled back to the modal split and possibly distribution model, and amended modal matrices returned to the assignment process. The properties of such cobweb methods to determine equilibria in markets were well established in the economics literature; in particular, it was known that simple feedback was extremely inefficient and the existence of convergence was dependent on the demand and cost elasticities. Nevertheless, application of and lip service to such feedback methods began to appear in the 1970s and 80s.

A few scholars, however, had already begun to investigate ways to combine the trip distribution, mode split and traffic assignment steps, and eventually rediscovered Beckmann's formulation. Perhaps the first to embark on this line of thinking was Murchland (1970), but his efforts were largely unsuccessful. Subsequently, Evans (1976) and Florian and Nguyen (1975, 1978) proposed formulations that were effectively special cases or elaborations of Beckmann's original formulation. More importantly, they proposed convergent solution algorithms. Upon evaluation (Boyce et al, 1988), only the algorithm of Evans proved to be practical for solving small problems, let alone problems of a realistic size.

Even with this advance in understanding, these combined models were largely a research curiosity and effectively unknown to practitioners. Initial efforts to implement such models occurred in the early 1980s and continued through the 1990s (Boyce et al, 1983; Safwat and Magnanti, 1988; Lam and Huang, 1992; Fernandez et al, 1994). These early combined, or integrated, models had serious limitations as compared with professional practice. Often they represented only a single class of travel or trip purpose. Generally, they were much smaller in scale than the models used in practice. And, the solution properties of the models were still only beginning to be understood.

5 Current Status of Integrated Models

During the past ten years, three multiclass, integrated models have been implemented and applied. The first multiclass integrated model was implemented by Lam and Huang (1992). In their model, however, classes correspond to modes, so this model is not comparable to the other models shown. The second model represents the work of de Cea et al (2003), and is arguably the most advanced model available today. The third example is a model for the Chicago Region, implemented at the same scale and detail used in professional practice, by Boyce and Bar-Gera (2003).

Each of these models was estimated from available travel surveys and validated against census data. The models were solved with research codes or commercial software systems. In the case of Santiago, Chile, the procedure has evolved to the status of commercially available software, called *ESTRAUS*. All three models are solved with algorithms that may be traced back to the contribution of Evans (1976). Subsequent to these developments, Bar-Gera and Boyce (2003) integrated the origin-based assignment algorithm of Bar-Gera (2002) into a single-class, integrated model. This algorithm may also have the potential to solve multiclass models more precisely and faster than methods presently available; for details, see Boyce and Bar-Gera (2004).

It is interesting and important to note the close correspondence between the Evans algorithm for solving an integrated model and the four-stage procedure, which is actually a primitive algorithm for solving an unstated, integrated model. Having noted this point, our current understanding of integrated models provides guidance for solving the four-stage procedure with feedback, as is now required in the US. The gist of this insight is to recognize that the trip tables and road link flows are the solution variables of the problem must be adjusted from iteration to iteration to drive the solution towards equilibrium.

References

H. Bar-Gera "Origin-based algorithm for the traffic assignment problem," *Transportation Science* 36, 398-417 (2002).

H. Bar-Gera and D. Boyce "Origin-based algorithms for combined travel forecasting models," *Transportation Research* 37B, 405-422 (2003).

M. Beckmann, C. B. McGuire and C. B. Winsten, *Studies in the Economics of Transportation*, Yale University Press, New Haven, 1956.

M. E. Ben-Akiva, "Structure of passenger travel demand models," *Transportation Research Record* 526, 26-42 (1974).

D. E. Boyce, K. S. Chon, Y. J. Lee, K. T. Lin and L. J. LeBlanc, "Implementation and computational issues for combined models of location, destination, mode and route choice," *Environment and Planning A* 15, 1219-1230 (1983).

D. E. Boyce, L. J. LeBlanc and K. S. Chon, "Network equilibrium models of urban location and travel choices: a retrospective survey," *Journal of Regional Science* 28, 159-183 (1988).

D. Boyce and H. Bar-Gera, "Validation of multiclass urban travel forecasting models combining origin-destination, mode, and route choices," *Journal of Regional Science* 43, 517-540 (2003).

D. Boyce and H. Bar-Gera, "Multiclass combined models for urban travel forecasting," *Network and Spatial Economics* 4, 115-124 (2004).

A. J. Daly and S. Zachary, "Improved multiple choice models," in *Determinants of Travel Choice*, D. A. Hensher and M. Q. Dalvi (eds), Saxon House, Westmead, 1978.

J. De Cea, and J. L. Fernandez, V. Dekock, and A. Soto, "ESTRAUS: a computer package for solving supply-demand equilibrium problems on multimodal urban transportation networks with multiple user classes," presented at the Transportation Research Board Annual Meeting, Washington, DC, 2003.

S. Erlander and N. F. Stewart, *The Gravity Model in Transportation Analysis*, VSP, Utrecht, 1990.

S. P. Evans, "Derivation and analysis of some models for combining trip distribution and assignment," *Transportation Research* 10, 37-57 (1976).

J. E. Fernandez, J. de Cea, M. Florian, E. Cabrera, "Network equilibrium models with combined modes," *Transportation Science* 28, 182-192 (1994).

M. Florian, S. Nguyen and J. Ferland, "On the combined distribution-assignment of traffic," *Transportation Science* 9, 43-53 (1975).

M. Florian and S. Nguyen, "A combined trip distribution, modal split and trip assignment model," *Transportation Research* 12, 241-246 (1978).

W. H. K. Lam and H.-J. Huang, “A combined trip distribution and assignment model for multiple user classes. *Transportation Research* 26B, 275-287 (1992).

M. L. Manheim, *Fundamentals of Transportation Systems Analysis,* MIT Press, Cambridge, MA, 1979.

B. V. Martin, F. W. Memmott, 3rd, and A. J. Bone, *Principles and Techniques of Predicting Future Demand for Urban Area Transportation*, Research Report No. 38, M.I.T., Cambridge, MA, 1961.

D. McFadden, “Conditional logit analysis of qualitative choice behavior,” in *Frontiers in Econometrics*, P. Zarembka (ed), Academic Press, New York, 1973.

D. McFadden, “Modeling the choice of residential location,” in *Spatial Interaction Theory and Planning Models*, A. Karlqvist et al (eds), 75-96, North-Holland, Amsterdam, 1978.

J. D. Murchland, “Road network traffic distribution in equilibrium,” in *Mathematical Models in the Social Sciences,* 8, Anton Hain Verlag, Meisenheim am Glan, Germany, 145-183, 1970 (in German translation).

J. de D. Ortúzar and L. Willumsen, *Modelling Transport*, 3rd Edition, Wiley, Chichester, 2001.

M. Patriksson, *The Traffic Assignment Problem: Models and Methods*, VSP, Utrecht, 1994.

K. N. A. Safwat and T. L. Magnanti, A combined trip generation, trip distribution, modal split, and trip assignment model,” *Transportation Science* 18, 14-30 (1988).

H. C. W. L. Williams, “On the formati5on of travel demand models and economic evaluation measures of user benefit,” *Environment and Planning A* 9, 285-344 (1977).

H. C. W. L. Williams and M. L. Senior, “Model-based transport policy assessment; Part II: Removing fundamental inconsistencies from the models,” *Traffic Engineering and Control* 18, 464-469 (1977)

A. G. Wilson, *Entropy in Urban and Regional Modelling*, Pion, London, 1970.

H. J. Wooton and G. W. Pick, “A model for trips generated by households,” *Journal of Transport Economics and Policy* 1, 137-153 (1967).