Disaggregate Behavioral Travel Demand's RUM Side
A 30-Year Retrospective

Daniel McFadden
Department of Economics, University of California, Berkeley
March 2000 (revised July 2000)

RUM *Adj.* (British Slang) Strange, odd, peculiar. E.g."That new kid's a bit rum, he's always trying to snatch the best bird." The Internet Dictionary of Slang, 2000

ABSTRACT

This resource paper is intended to give a historical account of the development of the methodology of disaggregate behavioral travel demand analysis and its connection to random utility maximization (RUM). It reviews the early development of the subject, and major methodological innovations over the past three decades in choice theory, data collection, and statistical tools. It concludes by identifying some topics and issues that deserve more work, and fearlessly forecasting the future course of research in the field.

ACKNOWLEDGMENT

The preparation of this paper was supported by the E. Morris Cox Endowment at the University of California, Berkeley. Discussions with Moshe Ben-Akiva, Tommy Garling, and Kenneth Train have been very helpful. This paper was prepared for presentation at the International Association of Travel Behavior Analysts, Brisbane, Australia, July 2, 2000, and at RU2000, Duke University, Aug. 6, 2000.

Disaggregate Behavioral Travel Demand's RUM Side
A 30-Year Retrospective

Daniel McFadden

1. Introduction

> *"Zones don't travel; people travel!"*
> Slogan, Travel Demand Forecasting Project, 1972

This resource paper gives a historical account of the development of the methodology of disaggregate behavioral travel demand analysis and its connection to *random utility maximization* (RUM). It discusses some current research topics, and concludes by identifying some issues that deserve more work. I am an occasional visitor rather than a permanent resident in the house of travel demand analysis. This has some advantages. Like a grandparent, I get to enjoy the growth of the baby without the daily drudgery, and distance may lend perspective. However, it also has disadvantages which will be evident in the course of this paper. I do not know all the residents or the issues between them, and will undoubtedly overlook both people and things that are important. Also, what I know best about the house is the baggage I bring with me. I will attribute more significance to this baggage than others might, and it will be up to the residents to redress the balance.

By my chronology, disaggregate behavioral travel demand analysis in the form that you now know it was born in 1970. Up through the 1960's, the dominant tool for travel demand analysis was the *gravity model*, which described aggregate traffic between origin and destination zones in terms of zone sizes and generalized travel cost; see J. Meyer and M. Straszheim (1970). This model was fairly successful in describing flows through the highway network, but was unable to forecast demand along some of the dimensions needed to evaluate major capital projects being planned through the 1960's. Specifically, gravity models could not easily handle issues of mode split, trip generation, and trip timing after introduction of new modes or other major system changes, and could not capture the impact on behavior of fine-grained changes in the transportation system such as bus route density or bus headways. In the years just before 1970, driven by the Northeast Corridor Project and the San Francisco Bay Area Rapid Transit Project, the U.S. Department of Transportation funded a number of research studies to develop new travel demand analysis methods that would be more flexible and more sensitive to factors controlled by transportation policy. A good deal of the initial research was done around Cambridge, Mass., at Charles River Associates (CRA) under the guidance of John Meyer, and at M.I.T. and Cambridge Systematics (CSI) under the guidance of Marvin Manheim and Paul Roberts. There were parallel and somewhat independent developments in England, notably in the work of M. Beesley, D. Quarmby, and A. Wilson.

Table 1. RUM Theory and Origins of the MNL Model	
1927	Thurstone introduces his *law of comparative judgment*, a model of imperfect discrimination in which alternative I with true stimulus level V_i is perceived with a normal error as $V_i + \varepsilon_i$. Thurstone shows that the probability $P_{\{i,j\}}(I)$ that alternative I is chosen over alternative j has a form that we now call binomial probit.
1950s	Marschak generalizes Thurstone's law of comparative judgment to stochastic utility maximization in multinomial choice sets, names it the RUM model, and analyzes the relationships between random utility functions and choice probabilities.
1959	Luce publishes an axiomatic treatment of choice behavior which postulates that the ratio of choice probabilities for I and j is the same for every choice set C that includes I and j; and calls this *Independence from Irrelevant Alternatives* (IIA). Luce shows that if this axiom holds, then one can associate with each alternative a positive "strict utility" such that choice probabilities are proportional to their strict utilities. Marschak proves for a finite set of alternatives that choice probabilities satisfying the IIA axiom are consistent with RUM.
1965	McFadden parameterizes the strict utilities in the Luce model in a form suitable for econometric applications, and calls this *conditional logit*, now known as multinomial logit (MNL) Results include a computer program for maximum likelihood estimation of the model and a simple demonstration of its consistency with RUM.
1960s	Binomial and multinomial logit models without the RUM connection are used by Warner, Theil, Nerlove, Wilson, Quarmby, and others, some with travel demand applications. Quandt introduces a random parameters travel demand model that is implicitly of RUM form.

My own involvement with travel demand analysis came through a Federal Highway Administration (FHA) research project on disaggregate modeling directed by Tom Domencich at CRA. Primary consultants on that project were the economists Peter Diamond and Robert Hall from M.I.T. From the economic principles of consumer demand, they developed a behavioral travel demand model that emphasized separable utility and multi-stage budgeting, so that the complex dimensions of trip generation, timing, destination, and mode choice could be broken into manageable segments, with "inclusive values" tying the segments together in a coherent utility-maximization framework. In the style of the times, they developed this theory for a representative utility-maximizing consumer, and were then faced with the problem of putting the model together with data on individual trips from trip diaries. I was visiting M.I.T. in 1970-71, and Diamond and Hall asked me if I could make their model empirically operational. I was recruited for this task because I had a machine in hand for analyzing discrete choice. In my graduate student days at the end of the 1950's, I had been very interested in axiomatic choice theories developed by mathematical psychologists, including the RUM model introduced by Thurstone and studied by Marschak, and Luce's 1959 theory of individual choice behavior. In 1965, I was asked by a Berkeley graduate student, Phoebe Cottingham, for suggestions on how to analyze her thesis data on freeway routing decisions by the

California Department of Transportation. I suggested developing an empirical model from Luce's choice axiom. Letting $P_{\mathbf{C}}(I)$ denote the probability that a subject confronted with a set of mutually exclusive and exhaustive alternatives **C** will choose I, Luce's axiom states that the ratio of choice probabilities for I and j was the same for every choice set **C** that included I and j; i.e.,

(1) $$P_{\mathbf{C}}(I)/P_{\mathbf{C}}(j) = P_{\{i,j\}}(I)/P_{\{i,j\}}(j).$$

Luce called the axiom written in this form *Independence from Irrelevant Alternatives* (IIA).[1] Luce showed that if this axiom holds, then one can associate with each alternative a positive "strict utility" w_i such that

(2) $$P_{\mathbf{C}}(I) = w_i / \sum_{k \in \mathbf{C}} w_k \ .$$

Taking the strict utility for alternative I to be a parametric exponential function of its attributes x_i, $w_i = \exp(x_i\beta)$, gave a practical statistical model for individual choice data. I named this the *conditional logit model* because it reduced to a logistic in the two-alternative case, and had a ratio form analogous to the form for conditional probabilities. I set about writing a computer program to produce maximum likelihood estimates for this model, a difficult exercise in the early days of FORTRAN when linear algebra and optimization routines had to be written from scratch. The program was finally finished in 1967, too late for Phoebe's thesis. However, I was then able to put her data through the conditional logit machine; see McFadden (1968, 1975).

Figure 1 summarizes the time line of development of RUM theory. The concept was originally put forward by Thurstone (1927) as a *law of comparative judgment* to describe imperfect discrimination in tasks such as choosing the loudest sound. It pictured an alternative I with true stimulus level V_i being perceived as $V_i + \varepsilon_i$, where the ε_i are independent normally distributed perception errors. Thurstone observed that the probability that I would be judged louder than j would satisfy $P_{\{i,j\}}(I) = \text{Prob}(V_i + \varepsilon_i \geq V_j + \varepsilon_j)$, and showed that this probability had what we now call a binomial probit form. Marschak (1960) and Block & Marschak (1960) generalized this model to stochastic utility maximization over multiple alternatives, and introduced it to economics. I believe Marschak is responsible for naming RUM. Marschak explored the testable implications of maximization of random preferences, and proved for a finite set of alternatives that choice probabilities satisfying Luce's IIA axiom were consistent with RUM. An extension of this result established that a necessary and sufficient condition for RUM with independent errors to satisfy the IIA axiom was that the ε_i be identically distributed with a Type I Extreme Value distribution, $\text{Prob}(\varepsilon_i \leq c) = \exp(-e^{-c/\sigma})$, where σ is a scale factor. The sufficiency was proved by A. Marley and reported by Luce and Supes (1965). I proved necessity, starting from the implication of the Luce axiom that multinomial choice between an object with strict utility w_1 and m objects with strict utilities w_2 matched binomial choice between an object with strict utility w_1 and an object with strict utility mw_2; see McFadden (1968, 1973). The reason that the extreme value distribution appears here is its *max-stable* property that the maximum of two independent extreme value random variables with the same

[1]Luce's axiom in its most general form states that for each $I \in \mathbf{A} \subseteq \mathbf{C}$, $P_{\mathbf{C}}(I) = P_{\mathbf{C}}(\mathbf{A}) \cdot P_{\mathbf{A}}(I)$; this allows probabilities to be either positive or zero. When they are positive, this condition is equivalent to (1). The strictly positive case is relevant for experimental or field data where one can never definitively rule out improbable events.

scale factor is again an extreme value random variable with this scale factor; see Domencich & McFadden (1975, p. 63).

I initially interpreted the conditional logit model as a model of a decision-making bureaucracy, with random elements coming from heterogeneity of tastes of various bureaucrats. It was then transparent that in an empirical model with data across decision-makers, the randomness in utility could come from both inter-personal and intra-personal variation in preferences, and from variations in the attributes of alternatives known to the decision-maker but not to the observer. This led me to emphasize in my 1973 paper on the conditional logit model the idea of an *extensive* margin for discrete decisions in contrast to the *intensive* margin that operates for a representative consumer making continuous decisions. I knew, from the work of John Chipman (1960), and from a review of Luce's book by my Berkeley colleague Gerard Debreu (1960), that Luce's axiom could give implausible results when groups of alternatives were similar in their unobserved characteristics. In my work with Domencich, we recast an example due to Debreu as the "red-bus, blue-bus" problem; see McFadden (1973), Domencich & McFadden (1975, p.77).

For the implementation of the Diamond-Hall model in Domencich's FHA project, I adapted my conditional logit machinery, and Domencich and I successfully estimated a disaggregate urban travel demand model that encompassed both work and shopping trips, mode choice, destination choice, and trip generation. This appeared as a Federal Highway Administration report in 1972, with excerpts in papers that I published in 1973 and 1974. The FHA report was eventually published as the book *Urban Travel Demand* by Tom Domencich and myself in 1975. The empirical travel demand model from this study is not much cited, but the key elements, random utility with the randomness coming from heterogeneity in tastes across individuals, separable preference structures that permit analysis of segments of complex decisions in which inclusive values are sufficient statistics for the opportunities within a branch, and estimation from disaggregated data on individual trips, have become everyday tools in travel demand analysis. The conditional logit model and its nested cousins are now sufficiently standardized so they can be estimated more or less mindlessly as options in many statistical software packages.

The 1970 CRA study is the one that I know best, but there were a number of prior and parallel research developments that deserve mention. The binomial logit model as a statistical tool had been employed, primarily in biostatistics, since the 1950's. I believe its first use in transportation was a study of mode choice by Stanley Warner in 1962. Studies of the value of travel time by Beesley (1965), Lisco (1967), and Lave (1970) applied discriminant, linearized probit, and binomial logit methods, respectively, to disaggregate mode split data, with results similar to those obtained later from the conditional logit model. The multinomial logit model was developed independently as a purely statistical model, without random utility underpinnings, by Theil (1969) and by A. Wilson (1972). In a paper done for the Northeast Corridor project, Richard Quandt (1970) explicitly introduced a random taste parameter demand model for mode choice. A number of papers in the late 1960's discussed the shortcomings of the gravity model and suggested more behavioral approaches; particularly noteworthy are Stopher & Lisco (1970), Hartgen & Tanner (1970), and Brand (1972). Thus, many of the elements of what we now call disaggregate behavioral travel demand analysis were "in the air" in 1970. My conditional logit machine provided a practical way to implement these ideas. Perhaps more importantly, it established the connection between random utility maximization as an organizing concept for model development, and the specification of empirical travel demand models.

Table 2. Application of RUM to Travel Demand Analysis	
1970	Domencich and McFadden estimate a travel demand system using separable utility and multi-stage budgeting, from Diamond and Hall, and McFadden's conditional logit model. Inclusive values to connect levels are calculated as probability weighted averages of systematic utility components at the next level down in the tree. The system includes work commute mode, and shopping generation, destination, and mode .
1972	Ben-Akiva develops the log sum formula for exact calculation of inclusive values
1977	McFadden, Williams, Daly & Zachary, and Ben-Akiva & Lerman develop independent RUM justifications for the nested MNL model. McFadden's analysis yields the GEV family of models. The Williams-Daly-Zachary analysis provides the foundation for derivation of RUM-consistent choice models from social surplus functions, and connects RUM-based models to Willingness-To-Pay (WTP) for projects.
1984	Ben-Akiva & McFadden develop multiple-indicator, multiple-cause (MIMC) models for combining revealed preference (RP) and stated preference (SP) data within a RUM framework. Morikawa develops this model further and applies it to intercity travel in the Netherlands. Extensive development and use of market research methods for collecting stated preference data are made by Hensher, Louviere, and others.
1989	McFadden introduces simulation methods that make it practical to estimate MNP and other open-form choice models.
1990s	The decade has seen extensive development and use of open and closed form choice models consistent with RUM, including GEV models and mixing in the parameters of MNL and nested MNL models. Economists and Psychologists have explored the cognitive foundations of RUM, and behavioral alternatives to RUM.

Table 2 gives a time line of RUM-related developments in travel demand analysis beginning with the model that Domencich and I implemented in 1971: There was one significant flaw in our analysis: While there was a random utility foundation for the ordinary conditional logit model, we did not establish that nested logit models had this foundation, or what form the random utility theory dictated for inclusive value terms. Our original study constructed inclusive values using linear averaging formulas brought over from the theory of separable preferences for a representative consumer. These turned out to be approximately right, but a superior exact formula that worked when nested MNL reduced to ordinary MNL models was found by Moshe Ben-Akiva in his 1972 Ph.D. thesis at MIT: An inclusive value at any level of the decision tree was given in this corrected formulation by a *log sum* formula, the log of the denominator of the choice model at the next level down the tree.

In the early days of disaggregate travel demand analysis, its value was questioned by many transportation policy analysts, who were skeptical that it could outperform gravity models. Research at M.I.T., Harvard, CRA, CSI, and Berkeley provided the push and the research results that led to its eventual acceptance. At Berkeley, I directed the Travel Demand Forecasting Project (TDFP), which set out to develop a comprehensive framework for transportation policy analysis using disaggregate behavioral tools.[2] TDFP used the introduction of BART as a natural experiment to test the ability of disaggregate travel demand models to forecast a new transportation mode. Table 3 below, taken from McFadden (1978), gives the work mode choice model that we estimated using data collected in 1972, before BART began operation, and subsequently used to predict BART patronage. The family annual income variable in this model enters as a linear spline with knots at $7.5K, $10.5K, and $15.5K; for comparison, median family income in 1972 was about $9.4K. The spline is zero at zero income, and is constant above $15.5K. The coefficients in the model give the slope of each segment of the spline. A number of features of this model are worthy of note. First, a major effort was made to measure household and alternative attributes accurately. Detailed questions were asked about wages and household income. Bus travel times were calculated for each commute trip from dispatcher bus schedules at the hour of the trip. Walk times were measured from actual distance on detailed maps, and took into account grade and number of streets to cross as well as length. Automobile routes were determined from a detailed highway network for the Bay Area, with link travel times adjusted to the hour of the commute, and adjusted further to take into account off-network distance.

The estimated model gives substantially higher values of travel time than are typical in value of time studies; see Lisco (1967) and Lave (1970). In particular, the value of auto in-vehicle time is very high, suggesting on its face that driving under commute conditions is much more onerous than working. Since measurement error in explanatory variables in general tends to attenuate coefficients, some part of our high values may be the result of more accurate measurement of travel attributes. They might also result from inaccurate measurement of the wage rate, which would tend to depress the magnitude of the cost coefficient and inflate the value of time estimate. One of our findings was that the estimated value of time was quite sensitive to specification of the independent variables; a simplified model with only time and cost variables gave values of on-vehicle time close to the 40 to 50 percent of wage that has typically been found elsewhere. This raises the possibility that customary values are biased downward by exclusion of some important explanatory variables. The model shows very little sensitivity to family income, with the spline showing no significant impact at any level. Since autos per driver is an explanatory variable, and income is likely to influence mode choice largely through auto ownership, this is not surprising. The pattern of interactions of number of drivers, number of automobiles, and commute mode indicates that adding an automobile for a fixed number of drivers tends to strongly increase all auto-using modes, particularly drive alone, at the expense of bus with walk access, while adding drivers without adding automobiles tends to increase use of bus with walk access.

---

[2]One of the contributions of TDFP was to guide a number of researchers into disaggregate travel demand analysis. Several of its research assistants, including David Brownstone, Steve Cosslett, Tim Hau, Ken Small, Kenneth Train, and Cliff Winston, went on to careers with a significant component in transportation analysis. Its research associates, Antti Talvitie and Chuck Manski, are also important contributors. On the methodological front, it developed methods for choice-based sampling and for simulation, and statistical methods for estimating and testing nested logit models, that laid the foundation for later results.

Table 3. Work Trip Mode Choice Model, Estimated Pre-BART

Mode	Description	Number	Percent
1	Auto Alone	429	55.6%
2	Bus with Walk Access	134	17.4%
3	Bus with Auto Access	30	3.9%
4	Car Pool	178	23.1%
	Total	771	

Independent Variable (Appearing for numbered modes)	Max. Likelihood Estimate	T-Statistic
Cost/Post-tax wage (Cents/Cents per min.) [1-4]	-0.0284	4.31
Auto In-Vehicle Time (min.) [1.3.4]	-0.0644	5.65
Transit In-Vehicle Time(min) [2,3]	-0.0259	2.94
Walk Time (min) [2,3]	-0.0689	5.28
Transfer Wait Time (min) [2,3]	-0.0538	2.30
Number of Transfers [2.3]	-0.1050	0.78
Headway of First Bus (min) [2,3]	-0.0318	3.18
Family Income (thousands of $ per year) [1]		
Effect up to $7.5K	-0.0045	0.05
Effect between $7.5K and $10.5K	-0.0572	0.43
Effect above $10.5K	-0.0543	0.91
Number of Persons in Hh who can drive		
Auto Alone Interaction [1]	-0.1020	4.81
Auto Access Interaction [3]	-0.9900	3.29
Car Pool Interaction [4]	-0.8720	4.25
Dummy if commuter is Hh head [1]	-0.6270	3.37
Employment Density at Work Location [1]	-0.0016	2.27
Home Location in (2) or near (1) CBD [1]	-0.5020	4.18
Autos per Driver, ceiling of one		
Auto Alone Interaction [1]	5.0000	9.65
Auto Access Interaction [3]	2.3300	2.74
Car Pool Interaction [4]	2.3800	5.28
Alternative-Specific Dummy		
Auto Alone Interaction [1]	-5.2600	5.93
Auto Access Interaction [3]	-5.4900	5.33
Car Pool Interaction [4]	-3.8400	6.36
Log Likelihood at Convergence	-1069.0	
Log likelihood with alternative dummies only	-844.3	
Values of Time as a Percent of Wage		
Auto In-Vehicle Time	226.8%	3.20
Transit In-Vehicle Time	91.2%	2.43
Walk Time	242.6%	3.10
Transfer Wait Time	189.4%	2.01

The critical experiment came when our pre-BART forecasts from 1972 data, adjusted for actual 1975 travel times and costs, were compared with the observed BART mode share in 1975. Table 4, taken from McFadden (1978), summarizes the results. The predictions in this table are made by summing over the representative sample the estimated choice probabilities for each alternative, with the new BART modes added to the list of alternatives in the MNL formula. Thus, the number in row I and column j is the sum, over all individuals in the sample actually choosing mode I, of the predicted probabilities that they would choose mode j. Diagonal elements in this table are then correct predictions. Coefficients of alternative-specific variables and interactions for the new BART alternatives were assumed to be identical to the corresponding coefficients for the existing bus variables. This was a weak point in new mode forecasting using pure revealed preference models, and these days one could probably do better by incorporating some stated preference information.

The standard errors on predicted shares are substantial, and reflect parameter uncertainty which does not average out over the sample. Row totals give observed post-BART mode splits in the sample, and column totals give predicted mode splits. Dividing by sample size gives the actual and predicted mode shares, respectively. The Percent Correct for a mode is the ratio of a diagonal element to the column total, and in total is the ratio of the sum of the diagonal elements to the sum of all table elements. The Success Index is the ratio of the percent correct obtained by the model to that which would be obtained by a "dummies only" model that assigned observed mode shares as the choice probabilities for every member of the sample. The larger this index, the better the model is predicting relative to "chance". It is possible for a success index to be less than one, as the forecasting model does not have the information on actual mode shares. Prediction Error is simply the difference in the predicted and observed mode shares.

The model forecast a total BART share of 6.4 percent in 1975, closer to the actual share of 6.2 percent than might reasonably have been expected given the sizes of standard errors. The model under predicted the Auto Alone mode, and substantially over predicted transit share, 21.3% versus 18.4%. This may be the result of the IIA property of the MNL model, which does not account for the possibility that heterogeneous tastes induced more switching within transit modes than switching away from Auto Alone. In analyzing the sources of prediction failures, we found that the model substantially over predicted demand for modes requiring walk access. This may reflect a failure of the model to capture the value of walk time correctly, perhaps because taste heterogeneities led those who liked walking to locate where walk access times were moderate, so that reduced walk times mattered only to a restricted class of commuters.

The TDFP experiment identified some more generic issues for disaggregate behavioral forecasting: (I) Accurate measurement of travel time and cost components at the individual level is critical. Measures taken from network models, or estimates taken from users, can show large and systematic biases which disrupt disaggregate models. (ii) Alternative-specific effects and interactions are difficult to handle in forecasting, particularly for new transportation alternatives. The solution, to replace such effects with more comprehensive generic measurements of attributes may require non-market data on perceptions and attitudes. (iii) The machinery required to supply realistic demographics, socioeconomic variables, and transportation system attributes in a forecast year can be as challenging to develop as the behavioral models themselves. (iv) It can be a difficult modeling exercise to translate the natural language of policy initiatives into quantitative changes in the attributes of the alternatives each individual faces.

Table 4. Prediction Success Table for Pre-BART Model and Post-BART Data

Actual Choices	Cell Counts						
	Predicted Choices						
	Auto Alone	Bus/Walk	Bus/Auto	BART/Bus	BART/Auto	Carpool	Total
1 Auto Alone	255.1	22.2	6.3	1.5	13.7	79.1	378
2 Bus/Walk	11.6	36.4	3.0	1.7	1.4	13.9	68
3 Bus/Auto	1.2	2.8	0.7	0.0	1.6	2.6	9
4 BART/Bus	0.9	1.9	0.1	1.4	0.3	1.4	6
5 BART/Auto	8.9	3.1	1.8	0.7	8.8	9.7	33
6 Carpool	74.7	12.4	3.3	1.4	7.5	37.7	137
Total	352.3	79	15.2	6.6	33.3	144.5	631
Predicted Share	55.8%	12.5%	2.4%	1.1%	5.3%	22.9%	
(Std. Error)	(11.4%)	(3.4%)	(1.4%)	(0.5%)	(2.4%)	(10.7%)	
Actual Share	59.9%	10.8%	1.4%	1.0%	5.2%	21.7%	
Percent Correct	72.4%	46.1%	4.5%	21.0%	26.5%	26.1%	53.9%
Success Index	1.3	3.69	1.88	21	5	1.14	1.28
Prediction Error	-4.1%	1.7%	1.0%	0.1%	1.2%	6.8%	

By the end of the 1970's, the U.S. Department of Transportation had changed its focus to the practical problems of managing transportation facilities, and research funding for travel demand analysis dwindled. Many of the people who had been working on these problems moved over to energy demand modeling, and later to similar modeling problems in telecommunications, health care, and the environment. The flame of disaggregate travel demand modeling was kept alive through the continued commitment of Moshe Ben-Akiva and his associates at M.I.T., and growing interest and use of these models outside the U.S. Some of the work in other applications has also found use in travel demand analysis. For example, research by Dubin & McFadden (1984) and Cowing & McFadden (1984) on energy demand clarified how to link discrete and continuous choice behavior, and developed simulation methods for estimation and policy analysis. This proved useful in studies of the demand for automobiles and their VKT or energy consumption by Berkovic (1985), Mannering & Winston (1985), and Train (1986). A research project on electricity demand by Ben-Akiva and myself in 1984 set out a latent variable modeling framework for combining revealed preference and conjoint analysis data; see McFadden (1986), McFadden & Morikawa (1986), Train, McFadden, Goett (1987), and Goett, McFadden, Woo (1988). Work on telecommunications demand by Ben-Akiva, McFadden , & Train (1987) and Atherton, Ben Akiva, McFadden, & Train (1990) developed methods for dealing with large choice sets and choice set elicitation.

2. Developments

"Always do right - this will gratify some and astonish the rest."

Mark Twain, 1885

What are the new and important developments since the birth of disaggregate behavioral travel demand analysis? The remainder of this paper tries to answer this question. The past three decades over which disaggregate travel demand modeling has developed more or less coincide with the period of development of cognitive psychology, the computer revolution, and the transformation of market research into a quantitative subject. Each of these fields has had a substantial impact on progress and possibilities in travel demand analysis. In particular, cognitive psychology has provided a clearer understanding of how decisions are made and how they are influenced by various factors. It has also pointed out the cognitive anomalies that distort survey and experimental data. Better computers and software, combined with new statistical methods that make good use of computer power, have removed some of the barriers to natural models of travel behavior. Market research and applied psychology have developed methods for collecting data on choice behavior in hypothetical situations, and illuminating the choice process through measurement of perceptions, attitudes, and motivation. When used properly, these tools can sharpen the information revealed by actual travel behavior. In the following three sections, I will discuss, in order, developments in choice theory, in data, and in statistical methods. While I have separated these topics for purposes of presentation, they have been closely intertwined in their development, often feeding back from one to the other.

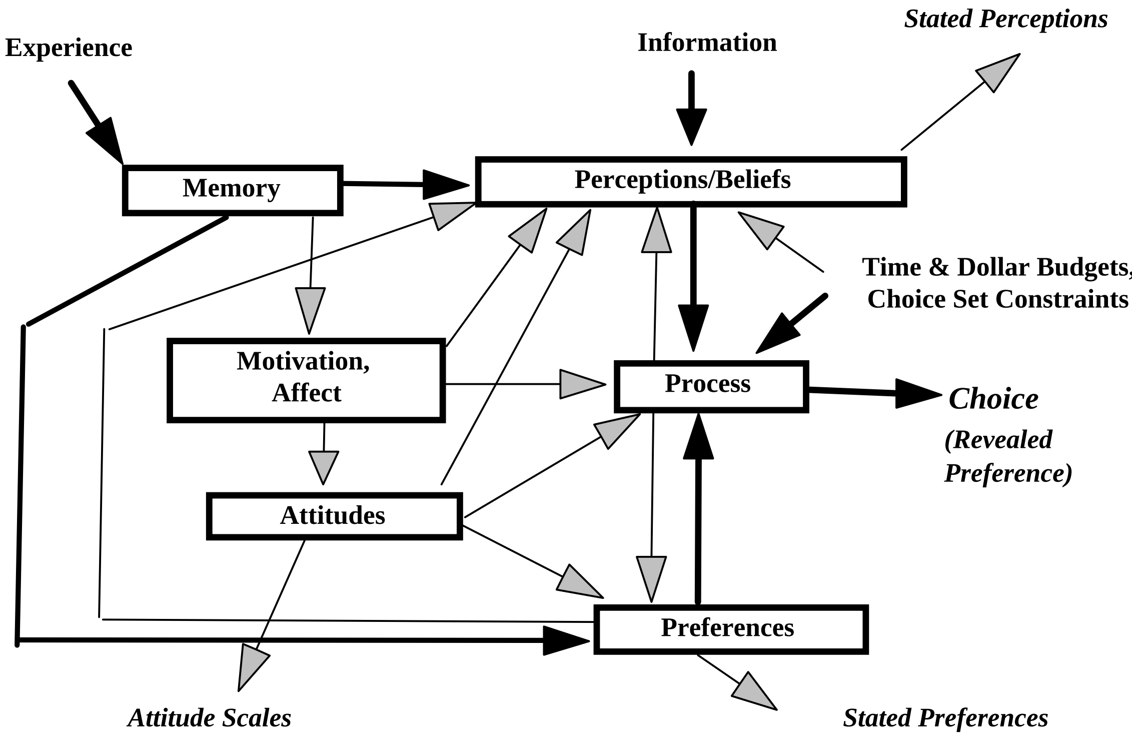
3. Choice Theory

"Economists have preferences; psychologists have attitudes."

Danny Kahneman, 1998

At a choice conference in Paris in 1998, a working group (Ben-Akiva *et al*, 1999) laid out the elements in a contemporary view of the theory of choice; an adaptation is shown in Figure 1. The figure describes one decision-making task in a lifelong sequence, with earlier information and choices operating through experience and memory to provide context for the current decision problem, and the results of this choice feeding forward to influence future decision problems. The heavy arrows in this figure coincide with the standard economic view of the choice process, in which individuals collect information on alternatives, use the rules of statistics to convert this information into perceived attributes, and then go through a cognitive process which can be represented as aggregating the perceived attribute levels into a stable one-dimensional utility index which is then maximized. The lighter arrows in the diagram correspond to psychological factors that enter decision-making, and the links between them. The concepts of *perception*, *preference*, and *process* appear in both economic and psychological views of decision-making, but with different views on how they work.

Figure 1.



Economic Consumer Theory and RUM

I will discuss the contribution of psychological elements to choice theory after first revisiting the connection of random utility models to the economic theory of choice. The heart of the standard economists' theory is the idea that consumers seek to maximize innate, stable preferences whose domain is the vector of attributes of the commodities they consume. The desirability of commodities will be determined by their attributes even if there are intermediate steps in which *raw goods* are transformed by the individual to produce *satisfactions* that are the proximate source of utility. e.g., commute trips are an input to employment, and employment is an input to the production of food and shelter.[3] An important implication of the theory is the *consumer sovereignty* property that preferences are predetermined in any choice situation, and do not depend on the alternatives available for choice. Succinctly, *desirability precedes availability*.

[3]The process of household production may yield structural restrictions that influence how different variables influence choice. It is possible to make the process of household production explicit; this is most readily done by considering *activity analysis* models of the decision-making unit. For some transportation applications, such as understanding the temporal organization of trips, chaining of trips, and the interplay of trips by household members, activity models are conceptually useful, and with sufficiently detailed data may be practical to estimate and use for forecasting. Users of activity models in transportation sometimes assume decision-makers follow heuristic rules that may be inconsistent with RUM. However, there is nothing intrinsic in activity analysis that requires this, and the suitability of RUM assumptions to characterize or approximate behavior in activity models involves the same issues of modeling strategy and empirical accuracy that appear in less disaggregated analysis of travel behavior.

The economists' theory has a vaguely biological flavor. Preferences are determined from a taste template coded in the genes, with experience and perceptions influencing how preferences consistent with the template are expressed. By taking a sufficiently relaxed view of how experience impacts perceptions and the unfolding of preferences, the economic theory of choice can be made to mimic psychological theories in which the individual is less organized, and more adaptive and imitative, in his choice behavior: choices gratify desires, even if desires are the result of a moment's whim, buffeted by emotion and experience. However, most applications of the economic theory leave out dependence of perception and preference on experience, and much of the power of the economists' approach lies in its ability to explain many behavior patterns *without* having to invoke experience or emotional state.

The original formulation of RUM as a behavioral model followed the economists' theory of consumer behavior, with features of the taste template that were heterogeneous across individuals and unknown to the analyst, as well as unobserved aspects of experience and of information on the attributes of alternatives, interpreted as random factors. If one then parameterized preferences and the distribution of the random factors, one ended up with a more or less tractable model for the probabilities of choice, expressed as functions of observed attributes of travel and individual characteristics. It is useful to review this derivation of the RUM explanation of travel behavior, taking a careful look at the meaning of its fundamental elements, and the scope and limitations of the models that come out. I believe this is particularly true for analysts who want to try to combine revealed preference and stated preference data, or who want to bring in cognitive and psychometric effects that are ignored in the economists' theory of choice.

In the economists' standard model, consumers have preferences over levels of consumption of goods and leisure, and seek to maximize these preferences subject to budget constraints for dollars and for time. One can start from the following setup: A consumer is assigned a travel portfolio with attributes (c,t,x,ξ), where c is travel cost, t is a vector of travel time components, x is a vector of other observed attributes of travel, and ξ is a vector of unobserved attributes of travel. The consumer has in addition an amount of pure leisure h and an amount of consumer goods g. The consumer's valuation of this allocation is given by a utility function $U^*(g,h+t(1-\lambda),x,\xi;s,\nu)$, where s is a vector of observed characteristics of the individual and ν is a vector of unobserved characteristics, with (s,ν) together determining tastes. The term $h+t(1-\lambda)$ appearing in this function is *total leisure*, made up of pure leisure h plus *effective* leisure contained in various travel time components. The idea is that a travel time component that is less onerous than working will have a coefficient λ that is between zero and one, interpreted as the portion of a travel hour in this component that is equivalent to work. Then, a travel time component that is equivalent to pure leisure will have $\lambda = 0$, a component that is exactly as onerous as work will have $\lambda = 1$, and a component that is more onerous than work will have $\lambda > 1$. The assumption that the leisure components of travel appear linearly and additively in the perception of total leisure is rather special, and not a necessary component of the standard economic model.[4] However, the linear additive form for total leisure is implicit in most travel demand models and allows some fruitful analysis of the structure of demand, so that I will impose it in what follows.

The utility function U^* can be interpreted as a partially reduced form that incorporates travel activities appearing as inputs to household production of satisfactions, but the consumer's problem of allocating time to employment in order to earn wage income to buy goods has not yet been solved.

---

[4]There is the suggestion from psychological experiments on time perception that mental accounting for time is non-linear and not necessarily time-consistent; see Lowenstein & Prelec (1992), Laibson (1997).

Then, attributes of the travel portfolio appearing in (x,ξ) appear as direct arguments in U^*. Travel cost c and non-leisure components of travel time t do not appear directly in the utility function, but rather through their impact on the consumer's dollar and time budget constraints,

(3) $$a + we = g + c,$$

(4) $$24 = h + e + t.$$

In these budgets, a is non-wage income and w is the wage rate, with both divided by an index of goods prices so that they have the same units as g, e is hours of employment, h is hours of pure leisure, and 24 is the daily budget of hours. Substitute g and h obtained by solving (3) and (4) into the utility function U^*, and define

(5) $$U(a\text{-}c,t\lambda,w,x,\xi;s,\nu) = \max_e U^*(a+we-c,24-e-t+t(1-\lambda),x,\xi;s,\nu).$$

This is an *indirect utility function*, and is the maximum of utility subject to the assigned travel portfolio and the constraints (3) and (4). It is increasing in a-c, decreasing in $t\lambda$, and quasi-convex in $(a\text{-}c,w)$. Conversely, any function with these properties is the solution of (5) for some utility function U^*. Then, specification of a RUM model can start from the random indirect utility function (5). Note that this function depends on the wage rate w and on net non-wage income after transportation cost, a-c, *not* on total income including the wage income component.[5] Also note that any monotone increasing transformation of an indirect utility function is an observationally equivalent indirect utility function, perhaps more transparent in structure.

The functional form of $U(a\text{-}c,t\lambda,w,x,\xi;s,\nu)$ will depend on the structure of preferences, including the trade-off between goods and leisure as a or w change, the role of household production in determining the structure of trade-offs between goods, and separability arising from tastes or from the organization of household production. Table 5 gives some alternative forms for the dependence of $U(a\text{-}c,t\lambda,w,x,\xi;s,\nu)$ on its arguments a-c, $t\lambda$, and w. The direct utility functions contain thresholds g_o and h_o that help explain the sensitivity of employment to wage, but do not enter the kernel of the indirect utility function. The formulas for the indirect utility functions omit additive terms that are uniform across travel portfolios and wash out of choice models. The dependence of these forms on x,ξ,s, and ν is suppressed, but these variables can enter the parameters of the forms in the table, and/or enter monotone increasing transformations of these forms; e.g., as additive terms. In the RUM analysis, discrete choice among travel portfolios is the result of each individual maximizing the indirect utility function $U(a\text{-}c,t\lambda,w,x,\xi;s,\nu)$ over a finite set of alternatives distinguished by their attributes (c,t,x,ξ). The leisure-linear form in Table 5 is the one most frequently encountered in the travel demand literature, and it and the goods-linear function can be treated as limiting cases of the Stone-Geary function. For a small data set, Train & McFadden (1978) analyze work mode split models of each of the forms in this table, and estimate θ to be about 0.7. Results obtained from study of the supply of labor suggest, if anything, a lower value for θ. Then the leisure-linear form encountered frequently in the travel demand literature may not be the best specification.

[5] For non-working individuals, w can be interpreted as an implicit opportunity cost of leisure foregone.

Table 5. RUM Functional Forms for Travel Demand		
Indirect Utility Function	**Direct Utility Function**	**Properties**
Leisure-Linear: $\alpha[(a\text{-}c)/w - t\lambda]$	$\log(g\text{-}g_o) + \alpha(h+t(1\text{-}\lambda))$	Zero income effect for goods, parameters α, g_o
Goods-Linear: $\alpha[(a\text{-}c - w{\cdot}t\lambda]$	$\alpha g + \log(h+t(1\text{-}\lambda)\text{-}h_o)$	Zero income effect for leisure, parameters α, h_o
Stone-Geary:: $\alpha w^{-\theta}[a\text{-}c - w{\cdot}t\lambda]$	$A(g\text{-}g_o)^{1-\theta}(h+t(1\text{-}\lambda)\text{-}h_o)^{\theta}$	Goods and leisure get fixed shares of uncommitted income, parameters A, θ, g_o, h_o, $\alpha = A\theta^{\theta}(1\text{-}\theta)^{1-\theta}$

All the models in Table 5 imply that the values of time in various travel activities are proportional to the wage rate, with the proportions λ, and that when utility is maximized across different travel profiles, higher wage individuals will tend to choose profiles that trade off added cost in order to save time. When combined with common specifications for unobserved components, the leisure-linear form implies that low-wage individuals show sharper discrimination among transportation alternatives based on time and cost differences than do high-wage individuals, with the reverse holding for the goods-linear model. The Stone-Geary form with θ = 0.7 implies that individuals become more sensitive to time differences and less sensitive to cost differences as wages increase.

Several aspects of the models in Table 5 deserve further comment. All of these models lead to indirect utility functions that are linear in non-wage income. When this is combined with an assumption that unobserverables enter *additively* with distributions that do not themselves depend on income, the result will be RUM choice models in which non-wage income washes out, even though socioeconomic status continues to enter through the wage rate and possibly through factors in s, such as education or occupation. A theorem first suggested by Williams (1977) and Daly-Zachary (1979) and formalized by McFadden (1981) establishes that choice behavior in this case can be described in terms of a “representative” consumer whose utility is the expected utility function and whose choice probabilities are given by derivatives of this function using Roy’s Identity. This is convenient both for the derivation of the choice probabilities and for applications of travel demand models where willingness-to-pay (WTP) for social policies such as transportation system improvements is needed. In economic terms, this is a case where Marshallian and Hicksian demand functions coincide and social aggregation of preferences is possible.

The assumption that travel costs enter the indirect utility function through the dollar budget, so that they always appear subtracted from non-wage income, is not special given the standard economist’s model, since in this setting “a dollar is a dollar” whose only role is as a medium of exchange. Psychologically, individuals may violate this dictum, keeping mental accounts in which dollars from different sources or for different uses are weighed differently. For example, some travel demand studies have found that out-of-pocket travel costs weigh more heavily in decisions than indirect or uncertain costs such as depreciation, maintenance, and expected accident costs. One could extend the models above to permit some cost components to enter with coefficients other than one to reflect their behavioral weight; i.e., c in the indirect utility function would be replaced by $c\kappa$ with c now a vector and κ behavioral weights. I will not introduce this generalization here, and note that while it is straightforward to implement in empirical travel demand analysis, it has some deep and troubling implications for economic analysis of travel demand models, such as determining WTP for social policy changes.

Next consider the unobserved factors (ξ,ν) in the indirect utility function $U(a\text{-}c,t\lambda,w,x,\xi;s,\nu)$. I will assume that ξ and ν are finite-dimensional. This assumption holds in most random utility models that are written down for applications, and imposes no loss of generality so long as the dimension of ν is at least as large as the size of the largest choice set to be considered. Consumer sovereignty requires that tastes be established *prior* to a specific choice problem; this implies that *the distribution of* ν *cannot depend on* $(a\text{-}c,t\lambda,w,x,\xi)$. In general, one might expect the distribution of ν to depend on s, say with a distribution function $H(\nu|s)$. Under mild regularity conditions one can represent ν in the form $\nu = h(p,s)$, where p is a vector with the dimension of ν whose components are independent uniformly distributed random variables on (0,1) and h is almost surely continuous in p, and then write the utility function as $U(a\text{-}c,t\lambda,w,x,\xi;s,h(p,s))$; see McFadden & Train (1998), Appendix Lemmas 2 and 3. Absorb this transformation into the definition of U and consider the random utility model $U(a\text{-}c,t\lambda,w,x,\xi;s,\nu)$ with ν uniformly distributed on a unit hypercube. Next, one might expect the distribution of ξ to depend in general on $t\lambda$ and x; e.g., unobserved comfort or security of a travel alternative may be correlated with observed variables such as travel time and the probability of getting a seat. As was the case for ν, ξ can be represented as a transformation of a uniformly distributed vector, and this transformation can be absorbed into the definition of U. It is likely that ξ does not depend on c, w, or s, because these are not intrinsic descriptors of a travel alternative and in principle could change without changing the remaining perceived attributes of the alternative. Also, if such dependence is present, it introduces a fundamental confounding of the causal economic effects of (c,w,s) and non-causal ecological effects. This makes the task of estimating causal travel demand models and using them for policy analysis virtually impossible. (Note however that it may be possible through market research methods that map the utility function of an individual to break the confounding by spurious ecological effects, and one then has to deal only with the difficult but not impossible task of tying stated preferences to real behavior.). For these reasons, I will assume that the conditional distribution of ξ given $(t\lambda,x,c,w,s)$ does not depend on (c,w,s), so that representing it as a transformation of a uniformly distributed vector does not alter the way c and w enter the indirect utility function, or its monotonicity and quasi-convexity properties in these variables. Then there is no essential loss of generality in writing the random utility function as $U(a\text{-}c,t\lambda,w,x;s,\varepsilon)$ with ε uniformly distributed on a unit hypercube, independently of the remaining arguments.[6] Finally, making a monotone transformation if necessary, we can assume that $U(a\text{-}c,t\lambda,w,x;s,\varepsilon)$ has all moments, or that its range is contained in the unit interval. This gives a canonical form that any RUM model for travel demand must satisfy. From here, one can get to applied models by specializing the specification of the function $U(\text{a-c,t,w,x;s},\varepsilon)$ in terms of its observed and unobserved variables.

In light of the preceding discussion, consider the original 1970 formulation of the RUM model for travel demand applications, which in the notation above might be written

[6]Technically, ε is a stochastic function, or random field, over the attributes (t,x) of the discrete alternatives, and with mild regularity conditions it can be assumed to be a continuous uniformly distributed random field.

$$(6) \qquad U(a\text{-}c,t\lambda,w,x;s,\varepsilon) = V(a\text{-}c,t\lambda,w,x,s) + \eta,$$

where $V(a\text{-}c,t\lambda,w,x,s)$ is a *systematic* or expected utility function and η summarizes the impact of all unobserved factors. In its implementation as a MNL or nested MNL model, the η were assumed to be independent or generalized extreme value distributed, independently of the arguments in the systematic utility function $V(a\text{-}c,t\lambda,w,x,s)$. In most applications, $V(a\text{-}c,t\lambda,w,x,s)$ was assumed to have a linear-in-parameters leisure-linear form,

$$(7) \qquad V(a\text{-}c,t\lambda,w,x,s) = \alpha[(a\text{-}c)/w - t\lambda] + z(x,s)\gamma,$$

where $z(x,s)$ was a vector of transformation of the observed attributes of alternatives and characteristics of the decision-maker. A natural question to ask in retrospect is how special is this specification, and to what degree can it be generalized to accommodate more general RUM-consistent behavior. The answer is that it is indeed quite specialized, no surprise given the strong IIA properties of the MNL model which also continue to hold within the lowest level nests in a nested MNL model. What is somewhat surprising is that a rather straightforward generalization of (6) and (7), to what are called mixed MNL models, can represent any well-behaved RUM-consistent behavior to detectable levels of accuracy. This result is established in McFadden & Train (2000); I will provide an intuitive summary.

Suppose choice behavior is consistent with a RUM model of the form $U(a\text{-}c,t\lambda,w,x;s,\varepsilon)$, where in light of the earlier argument we can assume that ε is uniformly distributed on a unit hypercube. Suppose the model is well-behaved in the sense that the arguments $(a\text{-}c,t\lambda,w,x;s)$ vary over a closed bounded domain, U is a continuous function of its arguments, the maximum number of alternatives in a choice set is bounded, and ε is a continuous random field over (t,x). Then, U has a uniform polynomial approximation in powers of ε and of transformations $Z\ (a\text{-}c,t\lambda,w,x;s)$ of the remaining arguments, uniformly for all possible alternatives in a choice set; this is just an application of the Weierstrauss theorem of elementary analysis. By collecting terms, we can write this approximation in the form

$$(8) \qquad U(a\text{-}c,t\lambda,w,x;s,\varepsilon) \approx \sum_{k=1}^{K} \alpha_k(\varepsilon)\cdot Z_k(a-c,w,t\lambda,x,s) \ ,$$

where the Z_k are transformations of the observed variables, and the α_k can be interpreted as random coefficients. The order K of the approximation can be chosen so that the probability that the approximation orders alternatives differently than the original is as small as we please. Further, we can scale the original and the approximation so that when the approximation is perturbed by adding an independent extreme value disturbance, the probability that the original and the perturbed approximation order the alternatives differently remains as small as we please. The choice probabilities implied by the perturbed approximation have a mixed MNL (MMNL) form,

$$(9) \qquad P_C(i) = \int \frac{e^{Z(a-c_i,w,t_i\lambda,x_i,s)\cdot\alpha}}{\sum_{j\in C} e^{Z(a-c_j,w,t_j\lambda,x_j,s)\cdot\alpha}} \cdot F(d\alpha) \ ,$$

where α and Z are vectors with components corresponding to the terms in (8), and F is the distribution of the α coefficients induced by the unobserved effects ε. From the construction, the distribution F does not depend on the attributes of alternatives. Because the perturbed approximation and the original order alternatives the same way except with very small probability, (9) will be as close as we please to the true RUM-consistent choice probabilities. It is immediate from its derivation that every MMNL model of the form (9) is RUM-consistent, provided the function (8) is an indirect utility function for each α in the domain of F. The model (9) has the interpretation of a MNL model of the usual linear-in-parameters form in which we allow the parameters to vary randomly, and in which we allow a flexible definition of the dependence of the "strict utility" of an alternative by introducing a sufficiently long series of transformations of the observed attributes of an alternative, interacted with observed characteristics of the decision-maker. If we wish, we can take F to be concentrated on a finite set of points, with the probability weights at these points treated as parameters. This is called a *latent class* model. If F is concentrated at a single point, MMNL reduces to MNL. Summarizing, I have outlined a result which says that any well-behaved RUM model can be approximated to any desired level of accuracy by a random-parameters or MMNL model, or more particularly by a latent class model, provided the transformations of observed variables and the random distributions that enter these forms are sufficiently flexible; see McFadden & Train (2000, Theorem 1). The MMNL model was first introduced by Cardell & Dunbar (1980), but was not widely used prior to the development of convenient simulation methods for estimation. Under the name *kernel logit*, it has been employed by McFadden (1989), Bolduc (1992), Brownstone & Train (1999), and Srinivasian & Mahmassani (2000) as a computational approximation to multinomial probit or as a general flexible RUM approximation. It is quite practical to apply MMNL models, and there is available good software for their estimation using simulation methods; see Revelt & Train (1998), Train (1999), Bhat (2000). Thus, for many purposes, this family is an attractive alternative to multinomial probit and other computationally demanding RUM models that are needed when ordinary MNL and nested MNL are not satisfactory.

The dimension of the integral in (9) is the dimension of the multivariate distribution of α. In many applications, this dimension will not need to be large to provide a quite flexible family that can approximate a wide range of RUM-consistent behaviors. It is often useful to think of the random elements as entering through a factor-analytic structure, with independent random elements loading with various weights in the coefficients of the $Z_k(a\text{-}c,t\lambda,w,x;s)$ variables. Then, the experience in other subjects that most multivariate data dependancies can be approximated well by a factor-analytic covariance structure with one to five factors is likely to also apply to travel demand models. The choice of independent extreme value perturbations in the approximation theory just outlined, leading to the mixed MNL form (8), was not essential, and other convenient perturbations such as independent normals could have been used instead, at a cost of adding one dimension to the integral in (8) and modifying the integrand to a form that might be less easy to evaluate numerically. However, I believe that for most applications, MMNL will prove to be the most convenient model.

It is useful to relate the approximation (8) to two approximation results that had previously appeared in the literature. In the early days, I gave a very elementary argument showing that any choice model, RUM-consistent or not, could be approximated to any desired degree of accuracy by a MNL model in which in general the attributes of all alternatives in the choice set could enter the "strict utility" of each alternative; see McFadden, Tye, and Train (1978). I called this the *mother logit* approximation, and suggested that it could be used as an alternative against which to test IIA. Because there was no easy way to tell whether a mother logit model was consistent with RUM, it did not provide a useful setup for estimating general RUM-consistent models or testing for RUM-consistency. In contrast, the mixed MNL models in (9) are guaranteed to be RUM-consistent if the linear forms

V(a-c,$t\lambda$,w,x,s) = Z(a-c,$t\lambda$,w,x;s)α $\equiv$ $\sum_{k=1}^{K} \alpha_k(\varepsilon)\cdot Z_k(a-c,w,t\lambda,x,s)$ satisfy the necessary and sufficient condition to be indirect utility functions: quasi-convex in (a-c,w) and increasing in a-c. The condition is easy to check when an adequate approximation can be attained with a-c appearing in a linear additive term, since then the monotonicity is satisfied if the coefficient on this term is positive and quasi-convexity is satisfied if the combination of terms involving w is convex. A sufficient condition for the latter property is that the coefficients of convex [resp., concave, linear] functions of w be positive [resp., negative, any sign]. More generally, a necessary and sufficient for quasi-convexity is

$$\text{(10)} \qquad \det\begin{bmatrix} \frac{\partial^2 Z(a-c,t\lambda,w,x;s)}{\partial(a-c)^2}\alpha & \frac{\partial^2 Z(a-c,t\lambda,w,x;s)}{\partial(a-c)\partial w}\alpha & \frac{\partial Z(a-c,t\lambda,w,x;s)}{\partial(a-c)}\alpha \\ \frac{\partial^2 Z(a-c,t\lambda,w,x;s)}{\partial w\partial(a-c)}\alpha & \frac{\partial^2 Z(a-c,t\lambda,w,x;s)}{\partial w^2}\alpha & \frac{\partial Z(a-c,t\lambda,w,x;s)}{\partial w}\alpha \\ \frac{\partial Z(a-c,t\lambda,w,x;s)}{\partial(a-c)}\alpha & \frac{\partial Z(a-c,t\lambda,w,x;s)}{\partial w}\alpha & 0 \end{bmatrix} \geq 0 \ .$$

This condition can be checked numerically over the range of the observations, or sufficient conditions can be imposed so that it holds globally; e.g., a sufficient condition for quasi-convexity is that each component of Z(a-c,$t\lambda$,w,x;s) is globally convex in (a-c,w) [resp., globally concave, linear] with a coefficient that is positive [resp., negative, unsigned]. The most common specification in travel demand applications, linear in (a-c,w), will automatically satisfy the quasi-convexity condition, so that the only questions regarding its consistency with true RUM-consistent behavior are whether it is increasing in a-c and approximates the observed choice behavior adequately. As discussed earlier, this commonly used linear-in-income approximation is special, with choice not directly influenced by non-wage income. Dependence on the wage rate w and on indicators of socioeconomic status s may mimic some of the effects of income, but this is nevertheless a possibly unrealistic and empirically refutable restriction on the choice model.

A recent approximation theorem has been obtained by Dagsvik (1994) in the context of modeling discrete choice behavior in continuous time. Dagsvik in effect starts from a RUM model with additive random effects, adds to it a very large common extreme value random variable so that all the observed choice behavior can be interpreted as the result a perturbation of this common extreme value effect, and then argues from some fundamental results on the existence of stochastic processes (specifically the de Haan representation of Poisson point processes) that there must exist a GEV stochastic process whose multivariate features, and in particular choice probabilities, approximately match those of the original stochastic process. Applied to the general RUM family analyzed above, Dagsvik's results establish that there exist GEV models that are dense in a class of well-behaved RUM models with additive disturbances. Then, the major difference in the Dagsvik result and the mixed MNL result (12) is that while neither is, strictly speaking, constructive, the latter is more readily adapted to practical approximation and computation.

Applications of travel demand models to transportation policy problems often call for estimation of WTP for policy changes. For example, a policy analysis may seek to determine the social benefit from introduction of a new transportation service, or modification of an existing service. When demand behavior is consistent with a RUM model with standard economic properties, WTP is given by conventional economic measures of consumer surplus. These measures are relatively simple to deduce from market demands (e.g., choice probabilities) when there are no income effects, or when the policy changes are sufficiently small so that linear corrections for income effects are accurate. To set ideas, suppose individuals have utility functions $U(a-c_i,t_i\lambda,w,x_i;s,\varepsilon)$ for alternatives i in a choice set **C**, and a policy initiative is considered which would change transportation attributes from (c_i',t_i',x_i') to (c_i'',t_i'',x_i''). The *compensating variation* or WTP is the adjustment in non-wage income necessary to give each individual the same level of utility after the initiative as was enjoyed before the initiative; i.e., the quantity y such that

$$(11) \qquad \max_{i\in\mathbf{C}} U(a-c_i',t_i'\lambda,w,x_i';s,\varepsilon) = \max_{i\in\mathbf{C}} U(a-y-c_i'',t_i''\lambda,w,x_i'';s,\varepsilon).$$

Note that y is a function of all the arguments in the utility function before and after the initiative, so that in particular it depends on ε. The mean WTP in a population of consumers is the expectation of their compensating variations with respect to the distribution of ε. One possible approach to estimating WTP is to simulate the computation in (11), calculating y from an estimated RUM model evaluated at a Monte Carlo draw of ε, and then averaging these calculated values over a large number of such draws. There are some useful bounds for this calculation. Define y_{ik} to satisfy

$$(12) \qquad U(a-c_i',t_i'\lambda,w,x_i';s,\varepsilon) = U(a-y_{ik}-c_k'',t_k''\lambda,w,x_k'';s,\varepsilon),$$

where i and k are any alternatives, not necessarily the chosen ones. Then, when m is the alternative that maximizes utility before the initiative, and n is the alternative that maximizes utility after the initiative, the compensating variation satisfies

$$(13) \qquad y_{mm} \geq \min_{k\in\mathbf{C}} y_{mk} = y = \max_{i\in\mathbf{C}} y_{in} \geq y_{nn}.$$

The intuition for these inequalities is that maximizing after the initiative reduces the compensating variation relative to what it would have been if the individual were forced to stay with m, and maximizing before the initiative raises the compensating variation relative to what it would have been if the individual had been forced to initially choose n. It is often relatively simple to calculate an average value $\bar{y}_{mm}'$ for the subpopulation that chooses alternative m before the initiative, and similarly an average value $\bar{y}_{nn}''$ for the subpopulation that chooses alternative n after the initiative. Then, population mean WTP satisfies the bounds

$$(14) \qquad \sum_{i\in\mathbf{C}} \bar{y}_{ii}'\cdot P_{\mathbf{C}}'(i) \;\geq \text{WTP} \geq\; \sum_{i\in\mathbf{C}} \bar{y}_{ii}''\cdot P_{\mathbf{C}}''(i) \;,$$

where $P_{\mathbf{C}}'(i)$ is a choice probability before the initiative and $P_{\mathbf{C}}''(i)$ is a choice probability after the initiative.

For many RUM models used in applications, computation of the bounds in (14) simplifies. For example, if random utility can be written as $V(a-c,t\lambda,w,x,s) + \eta$, with η containing all random effects, as in equation (6), then $\bar{y}_{ii}$ satisfies $V(a-c_i',t_i'\lambda,w,x_i',s) = V(a-\bar{y}_{ii}-c_i'',t_i''\lambda,w,x_i'',s)$ and is independent of the random effects. If, further, the systematic utility has a linear form, including an additive linear income term as in equation (7), so that $V(a\text{-}c,t\lambda,w,x,s) = \alpha[(a\text{-}c)/w - t\lambda] + z(x,s)\gamma$, then one obtains the explicit form

$$(15) \qquad \bar{y}_{ii} = w(z(x_i'',s) - z(x_i',s))\gamma/\alpha - w(t_i'' - t_i')\lambda - (c_i''\text{-}c_i').$$

Alternately, if $V(a\text{-}c,t\lambda,w,x,s) + \eta$ and the effect of the initiative is very small, one can make a Taylor's expansion of the post-initiative systematic utility about the pre-initiative values and drop higher-order terms to obtain the approximation

$$(16) \qquad \Delta\text{WTP} = \sum_{i\in \boldsymbol{C}} P_{\boldsymbol{C}}'(i)\cdot[\Delta x_i'\cdot z_x(x_i',s)\gamma/\alpha - \Delta t_i'\lambda\beta/\alpha - \Delta c_i'] \;,$$

where $\gamma = \partial V/\partial z$, $\beta = \partial V/\partial t\lambda$, and $\alpha = \partial V/\partial a$. Karlstrom (1998,2000) has developed an efficient method for computing WTP in the presence of income effects that can be interpreted as integration of this formula over a path from the pre-initiative to the post-initiative attributes.

Consider RUM models that are linear and additive in non-wage income, with a functional form $U(a\text{-}c_i,t_i\lambda,w,x_i;s,\varepsilon) = a - c_i + Q(t_i\lambda,w,x_i;s,\varepsilon)$. The Williams-Daly-Zachery theorem establishes that for this family the expected utility

$$(17) \qquad S_{\mathbf{C}}(\boldsymbol{c},\boldsymbol{t}\lambda,w,\boldsymbol{x};s) \equiv \mathbf{E}\max_{k\in\mathbf{C}} U(a\text{-}c_k,t_k\lambda,w,x_k;s,\varepsilon) = a + \mathbf{E}\max_{k\in\mathbf{C}}\{Q(t_i\lambda,w,x_i;s,\varepsilon) - c_k\}$$

behaves like an indirect utility function for a representative consumer, with $\boldsymbol{c}$ a column vector of travel costs for the alternatives in **C**, $\boldsymbol{t}$ an array of travel times with a row for each alternative, and $\boldsymbol{x}$ an array of other travel attributes, again with a row for each alternative. The choice probabilities equal the negatives of the derivatives of $S_{\mathbf{C}}(\boldsymbol{c},\boldsymbol{t}\lambda,w,\boldsymbol{x};s)$ with respect to the components of c. A further result is that WTP is given exactly by the difference in $S_{\mathbf{C}}(\boldsymbol{c},\boldsymbol{t}\lambda,w,\boldsymbol{x};s)$ after and before an innovation,

$$(18) \qquad \text{WTP} = S_{\mathbf{C}}(\boldsymbol{c}'',\boldsymbol{t}''\lambda,w,\boldsymbol{x}'';s) - S_{\mathbf{C}}(\boldsymbol{c}',\boldsymbol{t}'\lambda,w,\boldsymbol{x}';s).$$

If a MMNL form (9) with $Z(a\text{-}c,t\lambda,w,x;s)\alpha$ linear and additive in a with a (random) coefficient α_a provides an adequate approximation to RUM-consistent choice probabilities, then one also has as a good approximation

$$(19) \qquad S_{\mathbf{C}}(\boldsymbol{c},\boldsymbol{t}\lambda,w,\boldsymbol{x};s) = \boldsymbol{E}_\alpha \log\Big[\sum_{k\in \boldsymbol{C}} \exp(Z(a-c_k,w,t_k\lambda,x_k,s)\alpha)\Big]/\alpha_a \;.$$

Then a good approximation to WTP is obtained by starting from the familiar log sum formula for MNL and calculating its expectation with respect to α. Note however that (18) is valid, even as an approximation, only if a MMNL form that is linear and additive in a is a good approximation to the true RUM-consistent choice probabilities.

The original 1970 formulation of the RUM model was a special case of the derivation outlined in equations (3)-(19), but a number of useful elements have been added. First, the relationship between the functional form of the indirect utility function appearing in the RUM theory and consumer tastes for goods and leisure has been clarified and generalized. Second, practical methods have been developed for implementing RUM models without imposing restrictions such as the IIA axiom unless they are supported empirically, and the analysis has gone further to show that these methods place no effective restrictions on the class of RUM-consistent behavior that can be modeled. Third, the relationship between RUM-consistent choice probabilities and WTP has been established. Finally, RUM-consistent behavior has been placed in a context where the effects of experience and new information on perceptions and the expression of preferences can, in principle, be incorporated into the analysis. We shall see in the discussion to follow that this is insufficient to explain the cognitive process of making decisions, but may nevertheless be an appropriate and satisfactory tool for representing behavior within a system that can be used effectively for travel demand forecasting and transportation policy analysis.

The Psychology of Choice

Consider now the psychological elements that appeared in Figure 1, and the interpretation of perceptions, preferences, and process that come out of developmental, cognitive, and motivational psychology. I will give a brief description of the elements that are most significant for understanding choice behavior, and leave more detailed descriptions of theories and research findings to specialists in these fields. *Affect* and *motivation* are key concepts in a psychological view of decision-making, and determine *attitudes* which in turn strongly influence the cognition of the decision-making task. The lighter arrows in Figure 1 represent the linkages between these elements. Psychological observations and experiments establish that these factors have strong and sometimes surprising impacts on perceptions and on choice behavior. Many of the effects are stable and reliably reproducible in experiments. However, it is difficult to define and harness these psychological factors in a system that can forecast a broad spectrum of choice behavior.

Perceptions, attitudes, and preferences are conceptually different, but their definitions blur when one attempts to operationalize their measurement. Consider, for example, the four semantic differential (agree/disagree) questions in Table 6.

Table 6. Questions on Perceptions, Attitudes, and Preferences

1	At least once a month, there is an incident on Bus Line 7 that is a threat to personal security.
2	Personal security on public transit is a significant problem.
3	What I can save by taking the bus is not worth the risk to personal security.
4	I would pay $10 more in taxes each year to put security guards on public transit.

Question 1 elicits a perception, question 2 elicits an attitude that is primarily determined by perception, and question 3 elicits a preference that is strongly influenced by perception and/or attitude. The WTP question 4 elicits a preference, but the response could also be interpreted as determined by attitudes toward personal security and toward social responsibility. There is a clear difference between the first two questions, which involve no element of tastes, and the last two which do. However, all four questions reveal aspects of a constellation of beliefs, attitudes, and tastes that may not be fully differentiated within the individual, and may or may not be stable and consistent across time and decisions. The major scientific challenge to development of a psychological model of choice that can be used for travel demand applications is to find stable scales for attitudes, perceptions, and other psychological elements and establish that these scales can be used to forecast travel behavior more reliably than “reduced form” systems that map directly from experience and information to behavior.

The general view of psychologists is that behavior is highly adaptive and context-dependent. This is due in part to the dependence of perceptions on context, but more to the impact of emotional state (affect), motivation, and perceptions on the cognition of the task presented by a choice situation. Psychologists use the terms *problem-solving*, *reason-based*, or *rule-driven* to refer to behavioral processes that override utility-maximizing calculations, relying instead on principles or analogies to guide choice. Drazen Prelec (1991) distinguishes this view of decision-making from utility-maximization models by the cognitive processes involved:

> "Decision analysis, which codifies the rational model, views choice as a fundamentally technical problem of choosing the course of action that maximizes a unidimensional criterion, utility. The primary mental activity is the reduction of multiple attributes or dimensions to a single one, through specification of value trade-offs. For rule-governed action, the fundamental decision problem is the quasi-legal one of constructing a satisfying interpretation of the choice situation. The primary mental activity involved in this process is the exploration of analogies and distinctions between the current situation and other canonical choice situations in which a single rule or principle unambiguously applies. ... The purpose of rules must be derived from some weakness of our natural cost-benefit accounting system, and one might expect to find rules proliferating in exactly those choice domains where a natural utilitarianism does not produce satisfactory results."

Psychological elements add a fluidity and dynamic to decision-making; with many choices influenced by context, emotion, and adaptation. Most behavioral scientists would agree that these psychological elements are essential if one is to understand the *process* of decision-making, and probably necessary if one is to explain behavior in novel choice situations. There is no question that human choice behavior in a variety of laboratory settings shows striking deviations from the predictions of the economists' standard model, at least in its simpler and more rigid formulations. Further, the economists' concepts of a calculus of utility assessment and maximization rarely appear at a conscious level in self-reported decision protocols. Economic choice theory takes for granted that the consumer is *aware* that a decision is called for, and *prepared* to make a choice. In psychology, awareness and preparation are themselves substantive elements in the cognitive process. Cognitive theories of choice emphasize processes whose consistency with maximizing behavior is at best rough and ready, and the result of learning and conditioning rather than "rational" planning; see for example Svenson (1998) and McFadden (1999). There may be room within these psychological views of decision-making for the economists' taste template and "rational" expression of preferences, but some psychologists would argue that the process of choice is so completely controlled by immediate and context-dependent proximate factors that the existence or nature of a rational taste template is essentially irrelevant to understanding choice behavior. Both intuition and experimental evidence support the view that heuristic rules are the proximate drivers of most human behavior. The question remains as to whether rules themselves develop, from genetic templates or by learning and selection, in patterns that are broadly consistent with RUM postulates, so that RUM models can approximate behavior.

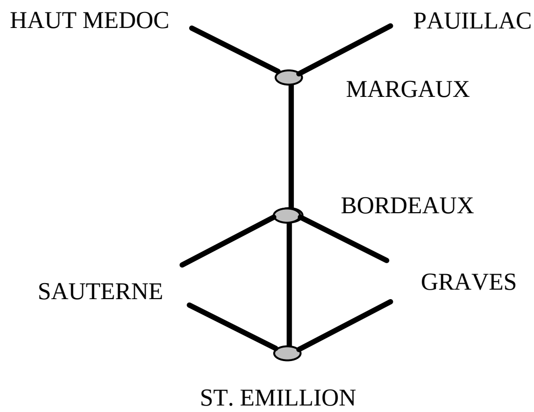
Even for routinized, "rational" decisions such as work trip mode choice which may be consistent with the economists' standard model, psychological elements are likely to be important in the construction and reinforcement of preferences. In addition, psychological and psychophysical measurements like perception and attitude scales can open windows to the decision-making process that are useful for model specification and estimation, even in applications where the economists' standard model does a good job of describing and forecasting behavior. The cognitive psychology of choice should be required study for all travel demand analysts, even the die-hard RUM modelers.

Market Research Methods and Travel Choices

In an assessment of the role of psychological elements in travel demand analysis done for purposes of evaluating transportation policies, it is unclear whether one *needs* to incorporate these elements in order to obtain reliable predictions of behavior for policy purposes, or for that matter whether we are capable of handling the resulting complexity when they are factored in. Economists and psychologists should recognize that what they consider the most interesting aspects of choice behavior are not necessarily important to transportation engineers. What is critical for transportation policy purposes is a "black box" that maps information about the transportation system into travel choices; the bottom line criterion is only that the black box work reliably. Every intervening construct within the box, such as an attitude, perception, or preference, is useful *only* if it is possible to provide both a mapping from information and experience to this construct *and* a mapping from this construct to choice that in tandem are more reliable than a direct "reduced form" mapping from experience and information to choice. Thus, the economists' standard model is useful for policy purposes *only* if the structure it imposes on the way information is processed enhances the reliability of policy models. Psychological elements need to meet this same standard if they are to be helpful for policy purposes. Some of the relativistic, context-dependent features of psychological theories of choice behavior which make them intuitive and natural for explaining one-off laboratory experiments militate against their easy use as mechanisms within a modeling system for policy-oriented travel demand analysis. Further, travel behavior may respond to transportation system attributes in the way that the economic theory predicts, even if the economists' standard model is wrong about how decisions are made. Thus, it is possible that psychology will provide a colorful and insightful language for describing travel behavior, but at the same time not add much to the reliability and explanatory power of policy forecasting models.

To emphasize this point, I adapt from McFadden (1999) an illustration of how we might go about understanding vision and forecasting visual perception in an applied task. Consider the simplified map of the wine-producing region around Bordeaux shown in Figure 2. Bordeaux appears to be closer to St. Emillion than to Margaux. However, the reader will immediately recognize that this is the classical Muller-Lyer optical illusion in which the distances are actually the same. Even after you are told this, St. Emillion looks closer. Could this illusion affect behavior? In fact, St. Emillion is more crowded than Margaux, perhaps due to other enophiles' illusions, but I doubt that anyone would claim that this is due to mass misreading of maps. We learn to be suspicious of our perceptions. We may see things cock-eyed, but we adopt conservative behavioral strategies that prevent us from deviating too far from our self-interest. One can learn a great deal about how visual cognition works by studying the breakdown regions where optical illusions occur, and draw from this lessons for how "normal" vision operates. Clearly a crude "what you see is what a camera sees" model of vision is false. Suppose now that you are trying to predict how people react to traffic signs when driving. Should you start from the library of optical illusions, or from the crude model of vision? I believe that it makes sense to start from a model that says that drivers will read what is written, and after that consult the library of optical illusions to see if what is written is likely to be misread. By this analogy, a standard RUM theory for travel choices may be a good starting point. One should then go on and ask whether psychological elements need to be added to explain travel demand behavior. Even if psychological factors are not in the end needed to explain some travel decisions, they can be extremely useful in specifying travel demand models. Understanding the cognitive and psychological elements in the decision-making process is particularly important when market research data are used to augment revealed choice data, because then one must be able to establish the links from market information through these psychometric variables to choice.

Figure 2. Roads in the Wine-Producing Region near Bordeaux



The most practical use of psychological concepts and measures in transportation demand analysis has come through market research. In studying demand for products, market researchers have confronted the problem that revealed market behavior may not provide a sufficient "natural experiment" to determine the mapping from experience and information into choice involving new products. Their remedy is to model explicitly the cognitive mechanisms that govern behavior, using psychometric data on consumer attitudes, perceptions, preferences, and intentions. These include multidimensional scaling of judgment data to obtain perception indicators, factor analysis of attitude inventories, stated preferences from hypothetical choice problems, and verbal protocols to describe the decision process. *Conjoint analysis*, the presentation of hypothetical choice tasks in an experimental design, has proven an effective way to elicit stated preference data.

One area where there the disagreements between the economic and psychological views of decision-making should be bridgeable is in the role of perceptions. There is nothing in the economic theory of consumer behavior that requires that perceptions be "rational", although economists believe that markets will often punish consumers who are predictably inconsistent, leading to selection against such belief systems. Many economists are comfortable with the idea that perceptions and attention are sensitive to context, and preference-maximizing behavior will emerge only in the "long-run" after experience has stabilized both perceptions and expressed preferences. There may be room here both for more careful attention by economists and psychologists to the formation and use of perceptions in decision-making, and a reexamination of the idea of a genetic taste template as an evolutionary adaptation. However, developments in these areas are unlikely to have much impact on the practice of behavioral transportation demand analysis in the near future.

4. Data

"Perception is not always reality."
Mercedez-Benz commercial, 1999.

The 1970's prescription for the data required for disaggregate travel demand analysis was to survey individuals on their travel behavior, through home and telephone interviews, and particularly through trip diaries. These data usually gave information on trips taken, and often the mode, travel time, and out-of-pocket cost for each trip segment. Less frequently, information was collected on route and the composition of the travelers (e.g., accompanied by children or not). These data usually did not give direct information on attributes of alternatives not selected; e.g., individuals taking auto were not asked about time and cost for a bus alternative, and individuals not taking a shopping trip during a particular period were not asked about travel possibilities during that period. To fill out the attributes of all the alternatives the individual faced, the typical procedure was to turn to road and transit networks, and use these networks to infer times and costs for all alternatives, including those chosen and those not chosen. There was often a gap between reported and calculated trip attributes, attributable to subject reporting error, network errors (e.g., due to uniform coding of travel times on secondary streets), and route assignment errors (e.g., consumers using non-optimal routes, or networks assigning routes that were in fact infeasible). When carefully collected, these methods gave fairly complete and accurate *revealed preference* (RP) data from real choice behavior.

There have been three major innovations in travel data collection since the 1970's. First, it was recognized that travel data could often be collected efficiently from on-board, screen line, or destination surveys. These have the advantage of concentrating observations on modes, locations, and times that are of particular interest for behavior and policy. In addition, they offer cost advantages over general home surveys. To illustrate, if one is interested in transportation system changes that will ease congestion caused by a new professional sports facility, an effective sample may intercept fans at the sports facility, and infer their travel behavior by comparing them to a control group of non-fans contacted at home. These are called *choice-based samples*, and the problem that they present is that they are sampling on the basis of the dependent variable, so that analysis must avoid confounding the legitimate effects of explanatory factors and the spurious effects of sampling. Section 5 reviews these methods.

A second innovation in data has had a major impact on travel demand analysis, and probably receives more attention than any other topic in travel demand research. This is the use of *stated preference* (SP) data, a shorthand for a variety of data that can be collected from individuals by offering them hypothetical choice tasks, eliciting attitudes and perceptions, and collecting subjective reports on preferences. Most of these variables and the methods used to measure them come from psychology via market research. In particular, conjoint analysis has proven that it can give a much more rounded view of the preferences of an individual than the one-dimension picture provided by revealed preference data.

The use of experiments rather than field surveys to collect data on demand has several major advantages. The environment of hypothetical choice can be precisely specified, with a design which allows straightforward identification of effects. Innovations in services can be studied. Large quantities of relevant data can be collected at moderate cost. However, as with any experiment, one can ask if laboratory behavior is a good predictor of field behavior. Good experimental technique can remove the most obvious sources of incongruity, but only field validation is fully convincing.

The important next step is to provide an analytic framework in which marketing data and revealed choice data can be combined to forecast the effects of transportation policy changes. In Section 5, I outline a latent variable model to handle these data that was developed by McFadden & Morikawa (1986) and Train, McFadden, and Goett (1987), and in various forms has been widely applied.

Numerous travel demand studies have now been published that use market research data. Some of the early applications are Morikawa (1989), Ben-Akiva & Morikawa (1990), Morikawa, Ben-Akiva, & Yamada (1991), Hensher & Bradley (1993), Louviere (1993, 1999), and Hensher, Louviere, & Swait (1989), and Brownstone & Train (1999). These studies show that carefully collected conjoint analysis data are on the whole measuring the same preferences as revealed preference data, with some calibration of location and scale required to adjust for perception and behavioral response differences between real and hypothetical choice situations. These studies appear to be quite successful in providing some structure to the distribution of tastes, and uncovering preferences along dimensions where RP data shows inadequate variation in attributes. I believe these techniques have progressed to the point that if one has the task of forecasting demand for a new travel alternative, it will probably be more reliable to establish the relationship between the attributes of existing and new alternatives by a SP experiment than to match new with existing alternatives as we did in the TDFP project in 1974.

It is my impression that data on perceptions have been collected less extensively or systematically than SP data on preferences. Although trip attributes collected in RP surveys are, perforce, perceptions that may vary from objectively measured attributes, and cognitive psychology has made us acutely aware of perceptual anomalies, I know of no study that has systematically examined the question of whether reported travel attributes are systematically biased, and what survey formats can be used to detect, control, and/or correct for such biases. There are lessons to be learned from study of survey methods in other fields; see for example Hurd, Merrill, and McFadden (1997).

A final major change in data for travel demand analysis comes from technologies for real time data collection. Starting with scanner data that can locate shopping destinations, through the use of GPS/cellular systems for tracking trip-makers, to the monitors for Intelligent Vehicle Systems, the possibilities are opening for observations on travel behavior at a level of fidelity and bandwidth that was unthinkable a decade ago. These data sources greatly expand the possibility of collecting *panel data* in which observations on repeated choices provide powerful "natural experiments" in which behavioral changes are coming primarily from changes in the choice setting rather than from changing tastes of the decision-maker. A major challenge in the immediate future will be to develop models and statistical methods that can handle these data. In marketing and other fields, this has often meant turning to nonparametric statistical methods that use linear filters suitable for streaming data, rather than more traditional parametric models. I expect soon to see wavelet or neural net models, calibrated using streams of data from trip-makers. The challenge to behavioral travel demand analysts will be to influence these acutely descriptive models so that they have some congruence with our understanding of how travel decisions are made, and the ability to produce the "what if" predictions needed for transportation policy analysis.

5. Statistical Methods

"What we see depends mainly on what we look for."
John Lubbock, 1952

The statistical problems presented by transportation demand applications recur throughout empirical choice theory and applied statistics more generally: the need to estimate model parameters that are embedded in nonlinear and not necessarily tractable forms, and the need for diagnostic tools to detect errors in specification and test hypotheses. Transportation demand analysis also needs systems for producing disaggregate and aggregate forecasts and policy scenarios that track statistical accuracy. Two problems that seem to be particularly irksome in transportation applications are identifying the choice set actually considered by the decision-maker, and dealing with the large number of possible travel portfolios when trip generation, destination, timing, mode, and route are all taken into account.

Applied RUM analysis, which is non-linear in parameters except in special cases, has generally used maximum likelihood estimation. This remains the workhorse for estimation of choice models, with some use of generalized method of moments estimators to incorporate market research data on stated perceptions and preferences. There have been incremental improvements in optimization algorithms, mostly buried in computer code and invisible to the applied researcher, but a few such as the E-M algorithm and Monte Carlo Markov Chain methods are available as named options. There have been obvious improvements in computation time and convenience. One useful development has been continued refinement in the theory of large sample hypothesis testing, and in particular the use of Lagrange Multiplier tests as a diagnostic tool. For choice model applications, this theory has been used to develop convenient tests of the IIA property of MNL models, and of random parameter dispersion in MMNL models. Perhaps the most significant innovations in statistical tools for transportation applications have come in sampling and in the use of simulation methods for estimation. In this section, I will review developments in the specification of RUM models, the theory of sampling, latent variable models for market research data, simulation-based estimation, and diagnostic tools.

RUM Families

An issue in the early disaggregate behavioral travel demand models using nested logit models was their consistency with RUM. I established this (for inclusive value coefficients between zero and one) in a 1978 paper in which I introduced a *Generalized Extreme Value* (GEV) family of models: Define a *GEV generating function* $H(w_1,...,w_J)$ to be a non-negative linear homogeneous function of $w \geq 0$ satisfying the properties that if any argument goes to $+\infty$, then H goes to $+\infty$; and the mixed partial derivatives of H exist, are continuous, and alternate in sign, with non-negative odd mixed derivatives. I showed that

$$F(\varepsilon_1,...,\varepsilon_J) = \exp(-H(e^{-\varepsilon_1},...,e^{-\varepsilon_J})) \quad , \tag{20}$$

is a joint distribution function whose one-dimensional marginals are then extreme value distributions. Consider a RUM model $u_i = V_i + \varepsilon_i$ for a set of alternatives $\mathbf{C} = \{1,...,J\}$, where the ε's have the distribution (20). Then one has

$$E \max_i u_i = \log(H(e^{V_1},...,e^{V_J})) + 0.57721 , \tag{21}$$

where 0.57721... is Euler's constant, and RUM choice probabilities are given by the logarithmic derivatives of (21), with the closed form

$$(22) \qquad P_{\mathbf{C}}(i) = e^{V_i} \cdot H_i(e^{V_1},...,e^{V_J})/H(e^{V_1},...,e^{V_J}) \; .$$

One example of a GEV generating function is the linear function $H = w_1 + ... + w_J$; this yields the ordinary MNL model. More complex GEV models, including nested logit, paired combinatorial logit, and cross-substitution logit, are obtained by repeated application of the following result: If sets $\mathbf{A}$,$\mathbf{B}$ satisfy $\mathbf{A}\cup\mathbf{B} = \mathbf{C}$, and $w_{\mathbf{A}}$, $w_{\mathbf{B}}$, and $w_{\mathbf{C}}$ are the corresponding subvectors of $(w_1,...,w_J)$, if $H^{\mathbf{A}}(w_{\mathbf{A}})$ and $H^{\mathbf{B}}(w_{\mathbf{B}})$ are GEV generating functions in $w_{\mathbf{A}}$ and $w_{\mathbf{B}}$, respectively, and if $s \geq 1$, then $H^{\mathbf{C}}(w_{\mathbf{C}}) = H^{\mathbf{A}}(w_{\mathbf{A}}{}^{s})^{1/s} + H^{\mathbf{B}}(w_{\mathbf{B}})$ is a GEV generating function in $w_{\mathbf{C}}$. The parameter 1/s is called an inclusive value coefficient. Discussions of useful specializations of the GEV family can be found in McFadden (1981), Small (1987), Bhat (1998), Papola (2000), and Fujiwara-Sugie-Moriyama (2000).

Somewhat different approaches that also established the consistency of nested MNL models with RUM were taken by Williams (1977), Daly & Zachary (1979) and Ben-Akiva & Lerman (1979). The Williams-Daly-Zachary formulation established two results that are useful more generally. First, they showed that an extreme value distributed random variable X can be written as the sum of two independent random variables Y and Z, with Z also extreme value distributed, if and only if the scale factor for X is at least as large as the scale factor for Z; a formal proof of this result is given in McFadden and Train (1998, Lemma 4). Second, they effectively showed that in the family of RUM models with an additive linear income term, expected maximum utility behaves like a "representative consumer" indirect utility function with the property that its price derivatives are proportional to the choice probabilities. A Nested MNL model with no income effects has the property that its choice probabilities are given by derivatives of its top level inclusive value, equation (5). Then, one can establish that a nested MNL model is consistent with RUM by showing, for suitable range restrictions on inclusive value coefficients, that its top level inclusive value meets the necessary and sufficient curvature conditions for an indirect utility function. This correspondence is established formally in McFadden (1981).

Probability Mixtures

An argument given earlier showed that a MMNL model is RUM-consistent if for each point in the support of the random parameters, the underlying MNL model is RUM-consistent. This proposition holds generally for mixing over *any* family of RUM-consistent choice models, with the mixing interpreted as taste heterogeneity in the corresponding population of RUM consumers. In particular, it is possible to consider mixtures over families of GEV models: If $H(w_1,...,w_J,\alpha)$ is a family of GEV generating functions with parameters α (which determine nesting structure, weights, and/or inclusive values), and F is a distribution over α, then the function $\exp(\int \log(H(e^{V_1},...,e^{V_J},\alpha))F(d\alpha))$ has logarithmic derivatives that are choice probabilities for a population whose behavior is described by this mixture of GEV models. When F has finite support, this function is a Cobb-Douglas combination of GEV models, with powers corresponding to the mixing probabilities. Since the MMNL model can approximate any RUM-consistent choice probabilities, it is unnecessary to consider mixtures of more complex models simply to attain a good approximation to RUM behavior. However, mixtures over models other than MNL may prove more parsimonious or more interpretable in some applications.

Sampling Methods

Problems in designing transportation surveys using only home-based random samples motivated the development of the theory of choice-based sampling. As indicated in Section 4, intercept, destination, and on-board surveys can provide information that is easier and less costly to obtain than home-based surveys. The question is then how to untangle by statistical analysis the choice behavior of the subjects from the patterns introduced by sampling. A solution has been provided by Manski & Lerman (1977), Manski & McFadden (1981), and Hsieh, Manski, & McFadden (1985). The setup of the estimation problem is that one has a vector of explanatory variables z and a choice i which are distributed in the target population with a density $p(z,i) \equiv P_C(i|z,\beta_o)p(z) \equiv Q(z|i,\beta_o)q(i)$, where $P_C(i|z,\beta_o)$ is the choice probability, defined as a member of a parametric family with true parameter vector β_o; $p(z)$ is the *marginal distribution* of the explanatory variables, $q(i) = q(i|\beta_o)$ is the *marginal distribution* of i, and $Q(z|i,\beta_o)$ is the *conditional distribution* of z given i, defined by Bayes law.

A simple random sample draws from the target population density $p(z,i)$. Exogenously stratified samples can be interpreted as draws of z from a density that may differ from the target population marginal density $p(z)$, followed by draws of i from the target population conditional density of i given z, $P_C(i|z,\beta_o)$. Then, the effects of exogenous sampling drop out of maximum likelihood estimation of the parameters β_o which appear only in the conditional density. Choice-based samples can be interpreted as draws of i from a density that may differ from the target population marginal density $q(i)$, followed by draws of z from the target population conditional density of z given i, $Q(z|i,\beta_o) = P_C(i|z,\beta_o)p(z)/q(i|\beta_o)$. This formula entangles the choice probability and the distribution of the explanatory variables, with the consequence that treating choice-based data *as if* it were random is generally statistically inconsistent; see Manski and Lerman (1977). One effective solution is to pool the observations from various strata, form for this pooled sample the conditional probability law of i given z, and then apply maximum likelihood to this conditional probability law.

I will describe this solution for general stratification schemes that include pure choice-based sampling as a special case. Suppose the data are collected from *strata* indexed s = 1,..., S. Each stratum is characterized by a sampling protocol that determines the segment of the population that qualifies for interviewing. Define $R(z,i,s)$ to be the *qualification probability* that a population member with characteristics (z,i) will qualify for the subpopulation from which the stratum s subsample will be drawn. For example, a choice-based stratum s from alternative 1 has $R(z,i,s) = 1$ if $i = 1$, and $R(z,i,s) = 0$ otherwise. The joint probability that a member of the target population will have variables (z,i) and will qualify for stratum s is $R(z,i,s)\cdot P_C(i|z,\beta_o)\cdot p(z)$. Then the proportion of the target population qualifying into stratum s, or *qualification factor*, is

$$r_s = \sum_z \sum_i R(z,i,s)\cdot P_C(i|z,\beta_o)\cdot p(z) . \qquad (23)$$

If f_s is the share of the sample in stratum s, then the conditional distribution of i given z and qualification into the pooled sample is

(24) $$\Pr(i|z,\beta_o) = \frac{\sum_{s=1}^{S} R(z,i,s)\cdot P_C(i|z,\beta_o)\cdot f_s/r_s}{\sum_{j=1}^{J}\ \sum_{s=1}^{S} R(z,j,s)\cdot P_C(j|z,\beta_o)\cdot f_s/r_s}\ .$$

When r_s is known or can be estimated from external data, the choice model parameters can be estimated consistently by applying maximum likelihood to these conditional probabilities; see the CML method of Manski & McFadden (1981). For choice-based samples in which qualification does not depend on z, this formula simplifies to

(25) $$\Pr(i|z,\beta_o) = \frac{P_C(i|z,\beta_o)\cdot e^{\alpha_i}}{\sum_{j=1}^{J}\ P_C(j|z,\beta_o)\cdot e^{\alpha_j}}\ ,$$

where $\alpha_i = \log(\ \sum_{s=1}^{S} R(z,i,s)\cdot f_s/r_s\ )$ can be treated as an alternative-specific constant. For MNL choice models, $\Pr(y|z,\beta_o)$ then reduces to a MNL formula with added alternative-specific constants. A more complicated example that arises in applications is *enriched sampling* in which a stratified exogenous sample is supplemented with a choice-based sample from alternatives that appear infrequently in the target population. To illustrate, take the case of a single exogenous sample that has $R(z,i,1) = 1$ for $z \in \mathbf{A}$ and zero otherwise, and a single choice-based sample that has $R(z,i,2) = 1$ for $i \in \mathbf{B}$ and zero otherwise. Then, $r_1 = p(\mathbf{A})$ is the share of the target population meeting condition $\mathbf{A}$, and $r_2 = q(\mathbf{B})$ is the frequency with which the target population makes a choice from $\mathbf{B}$. Then for $z \notin \mathbf{A}$, one has $\Pr(i|z,\beta_o) = P_C(i|z,\beta_o)/P_C(\mathbf{B}|z,\beta_o)$, and for $z \in \mathbf{A}$, one has $\Pr(i|z,\beta_o) = P_C(i|z,\beta_o)\cdot[f_1/r_1 + \mathbf{1}(i\in\mathbf{B})f_2/r_2]/[f_1/r_1 + P_C(B|z,\beta_o)f_2/r_2]$. These conditional probabilities used in the CML criterion permit consistent estimation of β_o. For forecasts and policy scenarios, one can weight each observed z from stratum 1 by r_1, and each observed z from stratum 2 by $r_2/P_C(\mathbf{B}|z,\beta_o)$ for $z \in \mathbf{A}$ and by $r_2/f_2\cdot P_C(\mathbf{B}|z,\beta_o)$ for $z \notin \mathbf{A}$, and then form the empirical expectation of the choice probabilities, before and after policy initiatives, with respect to this empirical baseline distribution for the explanatory variables.

The early literature on choice-based sampling dealt only with cross-section surveys and discrete dependent variables. These methods have been extended to problems where some components of the dependent variable are continuous, or the data comes from a panel in which subjects recruited in a choice-based survey design are then followed over time through subsequent choice situations; see for example McFadden (1997, 2000, Chap. 24).

Latent Variable Models for Market Research Data

The combined analysis of revealed preference and market research (stated preference) data requires statistical tools that allow for systematic and random differences. These may arise because of unobserved differences in real and hypothetical choice problems, including attributes of alternatives, the framing of the choice task, and the possibility for market discipline. I will summarize a model developed by Ben-Akiva and McFadden in 1984, and extended by Morikawa; see McFadden (1986), McFadden & Morikawa (1986). This is a specialization of the multiple-indicator, multiple-cause (MIMC) model that has been widely applied in psychology, sociology, and economics. Consider a binomial discrete response such as a mode choice, indexed by an indicator $d = \pm 1$. Suppose there is a latent random utility difference u^* explaining this response: $d = +1$ if and only if $u^* \geq 0$. The latent variable u^* is related by an equation $u^* = \alpha + \beta' y^* - \varepsilon$ to a $k\times 1$ vector of explanatory variables y^*, with an additive disturbance ε normalized so that $\mathbf{E}(\varepsilon|y^*) = 0$ and $\mathbf{E}(\varepsilon^2|y^*) = 1$. When y^* is observed without error, this is the standard latent variable model underlying RUM-consistent discrete response analysis. Now consider the more general possibility that y^* is a latent vector that is not observed directly, but that multiple indicators and/or multiple causes for y^* are observed. Specifically, let x denote a n×1 vector of observed indicators for y^*, z denote a m×1 vector of observed variables that are causal to y^*, and assume a MIMC model with normal disturbances,

(26) $$x = \gamma + Cy^* + \nu, \quad \text{and} \quad y^* = B(z-\mathbf{E}z) + \eta.$$

The n×k array C is a matrix of *factor loadings*, the k×m array B is a structural parameter matrix, and γ is a parameter vector. Both C and B may be subject to structural restrictions. Conditioned on z, the disturbances ν and η are assumed to have a multivariate normal distribution with zero means and covariances $\mathbf{E}\nu\nu' = \Lambda$, $\mathbf{E}\eta\eta' = \Gamma$, $\mathbf{E}\nu\eta' = 0$, and are assumed independent of ε. Then, given z, y^* is multivariate normal with mean zero and covariance matrix Γ, and x is multivariate normal with mean $\gamma + CB(z - \mathbf{E}z)$ and covariance matrix $\Lambda + C\Gamma C'$.

In a typical application such as mode choice, y^* is a vector of *perceived* attributes of the modes, x is a vector of psychometric ratings, and z is a vector of measured product attributes. For example, in a choice between automobile and bus, y^* may contain factors like "comfort" and "safety"; x may contain psychometric data such as ratings of autos on "steering responsiveness" and "smoothness of ride", attitude scales for the importance of various factors, or stated preferences; and z may contain variables like weight, horsepower, and braking distance, as well as consumer characteristics like income, age, and number of exposures to advertising. In general, the model (26) implies

(27) $$\mathbf{E}(y^*|z) = \gamma + CBz, \quad \text{and} \quad \mathbf{Cov}(x|z) = C'\Gamma C + \Lambda.$$

There are n(m+1)+n(n+1)/2 sample moments, sufficient to identify this many parameters between γ, C, B, Γ, and Λ unless there are deficiencies in rank. Then, at least n(k-m)+mk+k(k+1)/2 restrictions on the parameter arrays are necessary for identification. Provided the model is identified, the parameters in C, B, Λ, and Γ can be estimated by a LISREL or factor analysis program. These programs either minimize the distance between the sample moments and their population analogs in (3), using some distance metric, or else maximize the likelihood of the sample of observed x vectors, conditioned on z. An example of (26) is

(28) $$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ c_{41} & c_{42} & c_{43} \\ c_{51} & c_{52} & c_{53} \end{bmatrix} \begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \begin{bmatrix} z_1 - \boldsymbol{E}z_1 \\ z_2 - \boldsymbol{E}z_2 \end{bmatrix} + \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix},$$

where three unobserved factors with five indicators have factor loadings and other parameters that are identified by exclusion restrictions, the normalizations $c_{11} = c_{22} = c_{33} = 1$, and an assumption that the matrix Λ is block-diagonal between the first three indicators and the last two indicators. One can show constructively that the specified restrictions are sufficient for identification: Substitute the second set of equations in (26) into the first set, and estimate the resulting linear system by ordinary least squares. The regression on the first block of indicators identifies B. Given B, the regression on the second block of indicators identifies C. Given C and the assumption that Λ is block-diagonal between the first and second blocks, the covariance between the regression residuals in the first and second block identifies Γ. Given Γ and C, the within-block covariances identify Λ.

Let F(ε) denote the cumulative distribution function of the disturbance ε. Assume F has the symmetry property F(ε) = 1 - F(-ε), as is the case for normal or logistic disturbances. Then, $\mathbf{P}(d|y^*) = F(d(\alpha+\beta' y^*))$. If y^* is observed without error, then this is the standard binomial discrete response formula. For latent y^*, the probability law for the data is obtained by conditioning on x and z:

(29) $$\mathbf{P}(d|x,z) = \mathbf{E}\{\mathbf{P}(d|y^*|x,z\} \equiv \mathbf{E}\{F(d(\alpha+\beta' y^*))|x,z\}.$$

Given z, the vectors y^* and x are joint normal. Then, the conditional distribution of y^* given x, z is

(30) $$y^*|q \sim N(Rq,\Omega)$$

with

(31) $$R = \begin{bmatrix} \Gamma C'(C\Gamma C'+\Lambda)^{-1} \\ \Gamma C'(C\Gamma C'+\Lambda)^{-1}CB \end{bmatrix}, \quad q = \begin{bmatrix} x-\boldsymbol{E}x \\ z-\boldsymbol{E}z \end{bmatrix},$$

and

(32) $$\Omega = \Gamma - \Gamma C'(C\Gamma C'+\Lambda)^{-1}C\Gamma.$$

In particular, if ε is standard normal, corresponding to a probit model for the discrete choice, one has

(33) $$\mathbf{P}(d|y^*) = \Phi(d(\alpha+\beta' y^*)).$$

From (30), one has

(34) $$(\alpha+\beta' y^*-\varepsilon)|x,z \sim N(\alpha+\beta' Rq,1+\beta'\Omega\beta),$$

implying

(35) $$\mathbf{P}(d|q) = \Phi(d(\alpha+\beta' Rq)/[1+\beta'\Omega\beta]^{1/2}).$$

Given consistent estimates of R and Ω obtained from the LISREL analysis, one can estimate α and β in (35) by maximum likelihood. It is necessary to correct the usual asymptotic standard errors to account for the presence of estimated variables. The framework in equations (27)-(35) can be generalized to multivariate choice problems, but simulation methods may be required to make the estimation problem tractable.

In the model formulation above, the x indicators are treated as continuous. For stated preferences, this could correspond to rating data on alternatives. Alternately, in hypothetical choice problems, components of x may be latent utility levels for which there are discrete indicators for stated choices that maximize latent utility. This adds another layer of nonlinearity to the statistical analysis and produces multivariate normal integral expressions for conditional probabilities of observed RP and SP choices that in general will necessitate simulation-based estimators. How to maintain the tractability of the normal linear MIMC model structure without creating probabilities of essentially multinomal probit form that are difficult to work with is an open research question.

Computation and Simulation

From an era where estimation of a single multinomial logit model was a major time-consuming computational task, we have progressed to the point where simple multinomial logits are virtually instantaneous, even for massive numbers of alternatives and observations. This is nearly true for nested multinomial logit models, or logit models containing other non-linear elements, via general purpose maximum likelihood programs, although achieving and verifying convergence in such problems remains an art. What has remained a computationally hard problem is the evaluation of choice probabilities that cannot be expressed in closed form, but require numerical integration of moderately high dimension. For example, the multinomial probit model with an unrestricted covariance structure continues to resist conventional computation except for small problems.

Use of simulation methods has provided the most traction in obtaining practical representations and estimates for these computationally hard models. The first development of these methods for the multinomial probit model, by Manski & Lerman (1981), was followed by a paper of mine (McFadden, 1989) that clarified the statistical theory of estimation using simulation methods. This approach to estimation has benefitted from a great deal of research in the last decade on various practical simulators, including the use of Gibbs, Metropolis-Hastings, and other Monte Carlo Markov Chain samplers, use of pseudo-random and patterned random numbers such as Halton and Sobel sequences, and tools such as the simulated EM algorithm and the Method of Simulated Moments; see Bhat (2000), McFadden (1997), Train (1999). These methods have made it feasible to work with quite flexible models, such as multinomial probit and mixed multinomial logit and extreme value models. Considerable room for improvement in simulation methods remains. The problem of multidimensional numerical integration remains computationally hard, and this virtually guarantees that practical approximations based on a simulation of reasonable size are sometimes going to be quite poor. It is helpful in the design of simulation-based statistical inference to realize that there is an analogy between real data generated by a true data generation process and simulated data generated by a proposed model for this process, and statistical procedures that have good properties when applied to real data will often also have good properties when applied to simulated data, or to pooled real and simulated data. In particular, some of the statistical methods for dealing with measurement error and outliers in real data may prove useful for processing simulated data.

A model where simulation methods are usually needed, and relatively easy to apply, is the mixed MNL model. From the theory in Section 3, MMNL models can approximate any well-behaved RUM model, but their calculation requires specification and calculation of the expectation of MNL probabilities with respect to a probability distribution on the MNL model parameters. Because the MNL model itself is smooth in its parameters, the following procedure gives positive, unbiased, smooth simulators of the MMNL probabilities, and smooth simulators of their derivatives: Let α denote the vector of MNL parameters, and suppose α is given by a smooth parametric inverse mapping $\alpha = h(\varepsilon,\theta)$, where θ parameterizes the distribution of α and ε is uniformly distributed in a hypercube. This works easily for cases where the α are multivariate normal, or transformations of multivariate normals (e.g., log normal, truncated normal), and with somewhat more difficulty for other common distributions. The simulation procedure is then to draw a simulated sample of ε's, of size R, either at random or using some patterned random numbers such as Halton sequences, fix this sequence for all subsequent analysis, and treat the choice probabilities as if they are given exactly by the approximation

$$(36) \qquad P_{\mathbf{C}}(i) = \sum_{r=1}^{R} \frac{e^{Z(a-c_i,w,t_i\lambda,x_i,s)\cdot h(\varepsilon_r,\theta)}}{R\sum_{j\in \mathbf{C}} e^{Z(a-c_j,w,t_j\lambda,x_j,s)\cdot h(\varepsilon_r,\theta)}} \ .$$

A modest rate requirement on R, that it rise more rapidly than the square root of sample size, is sufficient to guarantee that either maximum likelihood or method of moments applied using the formula (36) will contain a negligible error arising from simulation in sufficiently large samples. To avoid misleading estimates of precision when sample sizes and R are moderate, one should use the statistical covariance formulas for possibly misspecified models; see McFadden & Train (1998). In applications where the inverse transformation $\alpha = h(\varepsilon,\theta)$ is not tractable, one can instead use importance sampling methods or a Metropolis-Hastings sampler to simulate (12).

Specification Testing: *IIA Tests*

The MNL model, with its IIA or "red-bus, blue bus" property, is a powerful tool for travel demand analysis when the IIA property is satisfied by an application, since it is easily estimated, allows drastic reduction of data collection and computation by sampling subsets of alternatives (see McFadden (1981, Atherton, Ben-Akiva, McFadden & Train, 1987), and gives an easy formula for forecasting demand for new alternatives. On the other hand, as the "red bus, blue bus" example illustrates, the model could produce seriously misleading forecasts if IIA fails. For this reason, there was an early interest in developing specification tests that could be used to detect failures of IIA. The first proposed test, due to McFadden, Tye, & Train (1978) and Hausman & McFadden (1984) required that one estimate the MNL model twice, once on a full set of alternatives **C**, and second on a specified subset of alternatives **A**, using the subsample with choices from this subset. If IIA holds, the two estimates should not be statistically different. If IIA fails, then there may be sharper discrimination within the subset **A**, so that the estimates from the second setup will be larger in magnitude than the estimates from the full set of alternatives. Let β_A denote the estimates obtained from the second setup, and Ω_A denote their estimated covariance matrix. Let β_C denote the estimates of the same parameters obtained from the full choice set, and Ω_C denote their estimated covariance matrix. (Some parameters that can be estimated from the full choice set may not be identified in the second setup, in which case β_C refers to estimates of the subvector of parameters that are identified in both setups.) The quadratic form $HM = (\beta_C - \beta_A)'(\Omega_A - \Omega_C)^{-1}(\beta_C - \beta_A)$ was shown by Hausman & McFadden to have a chi-square distribution when IIA is true. In calculating this test, one must be careful to restrict the comparison of parameters, dropping components as necessary, to get a non-singular array $\Omega_A - \Omega_C$. When this is done, the degrees of freedom of the chi-square test equals the rank of Ω_A - Ω_C. The simple form of the covariance matrix for the parameter difference arises because β_C is the efficient estimator for the problem.

A generalization of this test which is particularly easy to compute was proposed by McFadden (1987). Estimate the basic MNL model, using all the observations; and let $P_{\mathbf{C}}(i)$ denote the fitted model. Suppose **A** is a specified subset of alternatives. Create new variables in one of the following two forms for each observation:

(a) If x_i are the variables in the basic logit model, define new variables

$$\text{(37)} \qquad z_i = \begin{cases} x_i - \sum_{j\in\boldsymbol{A}} P_{\boldsymbol{A}}(j)\cdot x_j & \text{if } i \in \boldsymbol{A} \\ 0 & \text{if } i \notin \boldsymbol{A} \end{cases},$$

The variables z_i can be written in abbreviated form as $z_i = \delta_{i\mathbf{A}}(x_i - x_{\mathbf{A}})$, where $\delta_{iA} = 1$ iff $i \in \mathbf{A}$ and $x_A = \sum_{j\in A} P_{\mathbf{A}}(j)\cdot x_j$ and $P_{\mathbf{A}}(j)$ is calculated from the base model.

(b) Define the new variable

$$\text{(38)} \qquad z_i = \begin{cases} \log(P_{\boldsymbol{A}}(i) - \sum_{j\in\boldsymbol{A}} P_{\boldsymbol{A}}(j)\cdot\log(P_{\boldsymbol{A}}(j)) & \text{if } i \in \boldsymbol{A} \\ 0 & \text{if } i \notin \boldsymbol{A} \end{cases}$$

where $P_{\mathbf{A}}(j)$ is calculated from the basic model. A numerically equivalent form is obtained by first defining the systematic utility at the basic model estimated parameters,, $V_i = x_i\beta$, and then defining the new variable

$$\text{(39)} \qquad z_i = \begin{cases} V_i - \sum_{j\in\boldsymbol{A}} P_{\boldsymbol{A}}(j)\cdot V_j & \text{if } i \in \boldsymbol{A} \\ 0 & \text{if } i \notin \boldsymbol{A} \end{cases},$$

or more compactly, $z_i = \delta_{i\mathbf{A}}(V_i - V_{\mathbf{A}})$.

Estimate an expanded MNL model that contains the basic model variables plus the new variables z_i. Then test whether these added variables are significant. If there is a single added variable, as in the construction (b), then the T-statistic for this added variable is appropriate. More generally, one can form a likelihood ratio statistic

$$\text{(40)} \qquad \text{LR} = 2\left[\begin{pmatrix} \textit{Log Likelihood} \\ \textit{with } z's \end{pmatrix} - \begin{pmatrix} \textit{Log Likelihood} \\ \textit{without } z's \end{pmatrix}\right]$$

If IIA holds, this likelihood ratio statistic has a chi-square distribution with degrees of freedom equal to the number of added z variables (after eliminating any that are linearly dependent).

The test using variables of type (a) is statistically asymptotically equivalent to the Hausman-McFadden test for the subset of alternatives **A**. The test using variables of type (b) is equivalent to a one-degree-of-freedom Hausman-McFadden test focused in the direction determined by the parameters β. It will have greater power than the previous test if there is substantial variation in the V's across **A**. It is also asymptotically equivalent to a *score* or *Lagrange Multiplier* test of the basic MNL model against a *nested* MNL model in which subjects discriminate more sharply between alternatives within **A** than they do between alternatives that are not both in **A**. One minus the coefficient of the variable can be interpreted as a preliminary estimate of the inclusive value coefficient for the nest **A**. If there are subset-**A**-specific dummy variables in the basic model, then some components of the variables z_i of type (a) are linearly dependent on these original variables, and cannot be used in the testing procedure. Put another way, subset-**A**-specific dummy variables can mimic the effects of increased discrimination within **A** due to common unobserved components. One may get a rejection of the null hypothesis either if IIA is false, or if there is some other problem with the model specification, such as omitted variables or a failure of the logit form due, say, to asymmetry or to fat tails in the disturbances. Rejection of the IIA test will often occur when IIA is false, even if the nest **A** does not correctly represent the pattern of nesting. However, the test will typically have greatest power when **A** is a nest for which an IIA failure occurs.

The tests described above are for a *single* specified subset **A**. However, it is trivial to test the MNL model against *several nests at once*, simply by introducing an omitted variable for each suspected nest, and testing jointly that the coefficients of these omitted variables are zero. Alternative nests in the test can be overlapping and/or nested. The coefficients on the omitted variables and their T-statistics provide some guide to choice of nesting structure if the IIA hypothesis fails.

Specification Testing: Mixing in MNL Models

In light of the theoretical result in Section 3 that any well-behaved RUM model can be approximated by a mixed MNL model, satisfaction of the IIA property can be recast as a condition that there be no unobserved heterogeneity in the MNL model parameters. This suggests that a test for the validity of the IIA property, and specification test for the explanatory power to be added by introducing mixing, can be constructed using a Lagrange Multiplier approach. The advantage of this method is that the test procedure requires only estimation of basic MNL models, so that simulation-based estimators are not needed, and that it can test against a battery of alternatives at the same time. To describe the test, consider choice from a set **C**. Let x_i be a 1×K vector of attributes of alternative i. Suppose from a random sample n = 1,...,N one estimates the parameter α_e in the simple MNL model $L_{\mathbf{C}}(i;\mathbf{x},\alpha) = e^{x_i\alpha}/\sum_{j\in \mathbf{C}} e^{x_j\alpha}$, using maximum likelihood; constructs artificial variables for selected components t of x_i,

$$z_{ti} = (x_{ti} - x_{t\mathbf{C}})^2/2 \quad \text{with} \quad x_{tC} = \sum_{j\in \mathbf{C}} x_{tj}\cdot L_{\mathbf{C}}(j;\mathbf{x},\alpha_e)\,; \tag{41}$$

and then uses a Wald or Likelihood Ratio test for the hypothesis that the artificial variables z_{ti} should be omitted from the MNL model. This test is asymptotically equivalent to a Lagrange multiplier test of the hypothesis of no mixing against the alternative of a MMNL model $P_{\mathbf{C}}(i|\mathbf{x},\theta) = \int L_{\mathbf{C}}(i;\mathbf{x},h(\varepsilon,\theta))\cdot d\varepsilon$ with mixing in the selected components t of $\alpha = h(\varepsilon,\theta)$). The degrees of freedom equals the number of artificial variables z_{ti} that are linearly independent of x. McFadden & Train (1998) also generalize the preceding test so that an estimated MMNL model with some mixing components can be tested against the alternative that additional mixing components are needed.

6. A Look Ahead

Looking back over the past 30 years, how successful is disaggregate behavioral travel demand analysis based on the RUM hypothesis? I believe that it has shown itself to be capable of addressing a broad array of policy questions within a modeling framework that has generally promoted sensible models and helped avoid blind alleys. I believe that RUM analysis has been in net a good thing for travel demand modeling, despite the fact the dictates of tractability have sometimes led to implausibly restrictive models and abuse of data, as it has forced analysts to think consistently about how decision-makers respond to attributes of alternatives across a spectrum of travel decisions. Forecast accuracy is uneven, but I believe no more so that in most applied areas where complex phenomena are modeled with limited data and a fairly open field for model specification.

Some early possibilities have not yet been realized. The program to develop a unified comprehensive behavioral transportation demand system has not been completed, perhaps because such systems are now viewed as too unwieldy, or perhaps because funding agencies do not see the value. Where demand systems are used, in simulation models of transportation networks, they have tended to draw piecemeal from behavioral results and rely heavily on engineering calibration. This may be good enough for policy work, but it has not led to cumulative refinement of behavioral models and feedback between behavioral studies and policy outcomes. The statistical theory of simulation provides a framework in which engineering calibration can be systematized and integrated with behavioral studies, but this is not yet done in the transportation models I have seen.

The RUM foundation for travel demand models has been only lightly exploited. Models have generally conformed to the few basic qualitative constraints that RUM imposes, but have not gone beyond this to explore the structure of consumer preferences or the connections between travel behavior and other consumer demand behavior. The potentially important role of perceptions, ranging from classical psychophysical perception of attributes such as security and comfort, through psychological shaping of perceptions to reduce dissonance, to mental accounting for times and costs, remains largely unexplored.

What lies ahead for disaggregate behavioral travel demand analysis? Between the extreme arguments that psychological elements are on one hand essential to understanding choice behavior and on the other hand impossible to incorporate into transportation planning models, where does the future lie? I believe the answer is that the standard RUM model, based on a mildly altered version of the economists' standard theory of consumer behavior that allows more sensitivity of perceptions and preferences to experience, augmented with stated preference, perception, and attitude measures that uncover more of the process by which context molds choice, will increasingly become the dominant methodology for behavioral travel demand analysis. At present, the weak links in this setup are the lack of reliable scales for stated preferences, perceptions, and attitudes, and reliable mappings from experience and information to perceptions and attitudes. It would be useful to have a comprehensive research effort that identified the attitudes that are most relevant to travel behavior, and devised reliable methods for scaling these attitudes and relating them to experience.

Is the RUM model eventually doomed by the accumulation of psychological evidence that the cognitive process for decision-making is more complex and context-dependent, and by market research methods that measure the factors that RUM treats as random? I hope so. It would be disappointing if we cannot reach a deeper understanding and measurement of choice behavior than this model can provide. However, one should not rush to throw it out. RUM based on the economists' standard model may look rather old-fashioned at this point, and some might be tempted to cast about for a non-RUM framework as soon as an initial RUM specification works badly. My guess is that in fact for most travel demand applications, there is a RUM setup, perhaps enriched by some explicit structure to account for the formation and interaction of perceptions and attitudes, that will do a good job of representing behavior. This may not be a conventional MNL model with the usual measures of travel time and cost as explanatory variables, but I believe there is still a lot of room for travel demand analysts to develop richer and more realistic models of behavior within the paradigm of RUM, with consumers strongly motivated to maximize the desirability of perceived alternatives within a psychological context that may influence perceptions and tastes, and with modest extensions of traditional MNL functional forms to families like MMNL models.

References

Atherton, T., M. Ben-Akiva, D. McFadden, K. Train (1990) "Micro-simulation of Local Residential Telephone Demand Under Alternative Service Options and Rate Structures," in A. de Fontenay, M. Shugard, and D. Sibley (eds.), TELECOMMUNICATIONS DEMAND MODELLING, 137-163, Elsevier: Amsterdam.

Beesley, M. (1965) "The Value of Time Spent Travelling: Some New Evidence," ECONOMICA, 174-185.

Ben-Akiva, M. (1972) THE STRUCTURE OF TRAVEL DEMAND MODELS, Transportation Systems Division, Department of Civil Engineering, M.I.T., Ph.D. dissertation.

Ben-Akiva & Lerman (1979) "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility," in D. Hensher and P. Stopher, ed., BEHAVIOURAL TRAVEL MODELING, Coom Held: London.

Ben-Akiva, M. *et al* (1997) "Modeling Methods for Discrete Choice Analysis," MARKETING LETTERS, 8, 273-286.

Ben-Akiva, M. *et al*, (1999) "Extended Framework for Modeling Choice Behavior," MARKETING LETTERS, 10, 11-44.

Ben-Akiva, M. & S. Lerman (1985) DISCRETE CHOICE ANALYSIS, MIT Press: Cambridge MA.

Ben-Akiva, M., D. McFadden, & K. Train (1987) "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," THE RAND JOURNAL OF ECONOMICS, 18, 109-123.

Ben-Akiva. M. & T. Morikawa (1990) "Estimation of Switching Models from Revealed Preferences and Stated Intentions," TRANSPORTATION RESEARCH, 24A, 485-495.

Berkovic, J. (1985), "New Car Sales and Used Car Stocks: A Model of the Automobile Market,"RAND JOURNAL OF ECONOMICS; 16, 195-214.

Bhat, C. (1998) "Accommodating Flexible Substitution Patterns in Multidimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice," Transportation Research 32B, 425-440.

Bhat, C. (2000) "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," Working Paper, Department of Civil Engineering, University of Texas, Austin.

Block, H. & J. Marschak (1960) "Random Orderings and Stochastic Theories of Response," in I. Olkin, ed., CONTRIBUTIONS TO PROBABILITY AND STATISTICS, Stanford Univ. Press, Stanford.

Bolduc, D. (1992) "Generalized Autoregressive Errors in the Multinomial Probit Model," Transportation Research-B, 26B.

Brand, D. (1972) "The State of the Art of Travel Demand Forecasting: A Critical Review," Graduate School of Design, Harvard University.

Brownstone, D. & K. Train (1999) "Forecasting New Product Penetration with Flexible Substitution Patterns," JOURNAL OF ECONOMETRICS, 89, 109-129.

Cardell, N; Dunbar, F. (1980) "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research*, **14A**, No. 5-6, 423-434.

Chipman, J. (1960) "The Foundations of Utility," ECONOMETRICA, 28, 193-224.

Cottingham, P. (1966) MEASUREMENT OF NON-USER BENEFITS, Ph.D. dissertation, Dept. of Economics, Univ. of California, Berkeley.

Cowing, T. & D. McFadden (1984) MICROECONOMIC MODELING AND POLICY ANALYSIS, Academic Press: New York.

Dagsvik, J. (1994) "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Alternatives," ECONOMETRICA, 62, 1179-1205.

Daly, A. (1979) "Some Developments in Transport Demand Modeling," in D. Hensher and P. Stopher, eds., BEHAVIOURAL DEMAND MODELING, Croom Helm: London, 319-333.

Daly, A. & S. Zachary (1979) "Improved Multiple Choice Models," in D. Hensher & Q. Dalvi, eds., IDENTIFYING AND MEASURING THE DETERMINANTS OF MODE CHOICE, Teakfield: London.

Debreu, G. (1960) "Review of R.D. Luce *Individual Choice Behavior*," AMERICAN ECONOMIC REVIEW, 50, 186-188.

Domencich, T. and D. McFadden (1975) URBAN TRAVEL DEMAND, North Holland: Amsterdam.

Dubin, J. & D. McFadden (1984) "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," ECONOMETRICA; 52, 345-62.

Fujiwara, A., Y. Sugie, M. Moriyama (2000) "Nested Paired Combinatorial Logit Model to Analyze Recreational Touring Behavior," Working Paper, Hiroshima University.

Goett, A.; D. McFadden; C. Woo (1988) "Estimating Household Value of Electrical Service Reliability with Market Research Data," ENERGY-JOURNAL; 9, 105-20...

Hajivassiliou, V. & P. Ruud (1994) "Classical Estimation Methods for LDV Models using Simulation," in R. Engle & D. McFadden, eds HANDBOOK OF ECONOMETRICS IV, 2384-2441, North Holland: Amsterdam.

Hajivassiliou, V. & D. McFadden (1998) "The Method of Simulated Scores with Application to Models of External Debt Crises," ECONOMETRICA, 66, 863-896.

Hajivassiliou, V., D. McFadden, & P. Ruud (1996) "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," JOURNAL OF ECONOMETRICS, 72, 85-134.

Hausman, J. & D. McFadden (1984) "Specification Tests for the Multinomial Logit Model," ECONOMETRICA, 52, 1219-1240.

Hartgen, D. & G. Tanner (1970) “Behavioral Model of Mode Choice: Preliminary Report,” New York State Department of Transportation.
Hensher, D. (1994) “Stated Preference Analysis of Travel Choices,” TRANSPORTATION, 21, 107-133.
Hensher, D. & M. Bradley (1993) “Using Stated Response Data to Enrich Revealed Preference Discrete Choice Models,” MARKETING LETTERS, 4, 39-152.
Hensher, D., J. Louviere, & J. Swait (1999) “Combining Sources of Preference Data,” JOURNAL OF ECONOMETRICS, 87, 197-221.
Hoch, S. & G. Lowenstein (1991) “Time-Inconsisitent Preferences and Consumer Self-Control,” JOURNAL OF CONSUMER RESEARCH, 17, 492-507.
Hsieh, D., C. Manski & D. McFadden (1985) “ Estimation of Response Probabilities from Augmented Retrospective Observations,” JOURNAL-OF-THE-AMERICAN-STATISTICAL-ASSOCIATION; 80, 651-62..
Hurd, M., A. Merrill, & D. McFadden (1997) “Consumption and Savings Balances of the Elderly: Experimental Evidence on Survey Response,” in D. Wise, ed., FRONTIERS IN THE ECONOMICS OF AGING, 259-305, Univ/ of Chicago Press: Chicago.
Laibson, D. (1997) “Golden Eggs and Hyperbolic Discounting,” QUARTERLY JOURNAL OF ECONOMICS, 112, 443-477.
Lave, C. (1970) “The Demand for Urban Mass Transit,” REVIEW OF ECONOMICS AND STATISTICS, 52, 320-323.
Karlstrom, A. (1998) “Hicksian Welfare Measures in a Nonlinear random Utility Framework,” Working Paper, Department of Infrastructure and Planning, Royal Institute of Technology, Stockholm.
Karlstrom, A. (2000) “Non-linear Value Functions in Random Utility Econometrics,” IATBR Conference, Australia.
Lisco, T. (1967) THE VALUE OF COMMUTERS' TRAVEL TIME: A STUDY IN URBAN TRANSPORTATION, University of Chicago, Ph.D. dissertation.
Louviere, J. , M. Fox, & W. Moore (1993) “Cross-Task Validity Comparisons of Stated Preference Models,” MARKETING LETTERS, 4, 205-213.
Louviere, J. *et al* (1999) “Combining Sources of Preference Data for Modeling Complex Decision Processes,” MARKETING LETTERS, 10, 45-75.
Lowenstein, G. & D. Prelec (1992) “Anomalies in Intertemporal Choice: Evidence and An Interpretation,” QUARTERLY JOURNAL OF ECONOMICS, 107, 573-597.
Luce, R. D. (1959) INDIVIDUAL CHOICE BEHAVIOR, Wiley: New York.
Luce, R. D. & P. Supes (1965) “Preference, Utility, and Subjective Probability,” in R. Luce, R. Bush & E. Galanter, eds., HANDBOOK OF MATHEMATICAL PSYCHOLOGY, Wiley: New York.
Mannering, F. and C. Winston (1985) “A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization,” RAND JOURNAL OF ECONOMICS, 16, 215-36.
Manski,C. & S. Lerman (1977), “The Estimation of Choice Probabilities from Choice Based Samples,” ECONOMETRICA; 45, 1977-88..
Manski, C. & D. McFadden (1981) "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in C.F. Manski and D. McFadden (eds.), STRUCTURAL ANALYSIS OF DISCRETE DATA WITH ECONOMETRIC APPLICATIONS. 2-50, MIT Press: Cambridge..
Manski, C. & S. Lerman (1981) “On the Use of Simulated Frequencies to Approximate Choice Probabilities,” in C.F. Manski and D. McFadden (eds.), STRUCTURAL ANALYSIS OF DISCRETE DATA WITH ECONOMETRIC APPLICATIONS. 305-319, MIT Press: Cambridge.
Marschak, J. (1960) “Binary Choice Constraints on Random Utility Indicators,” K. Arrow, ed., STANFORD SYMPOSIUM ON MATHEMATICAL METHODS IN THE SOCIAL SCIENCES, Stanford Univ. Press: Stanford.
McFadden, D. (1968) “The Revealed Preferences of a Public Bureaucracy,” Dept. of Economics, Univ. of California, Berkeley.
McFadden, D. (1973) “Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), FRONTIERS IN ECONOMETRICS, 105-142, Academic Press: New York.
McFadden, D. (1974) "The Measurement of Urban Travel Demand," JOURNAL OF PUBLIC ECONOMICS, 3, 303-328.
McFadden, D. (1976) "The Revealed Preferences of a Government Bureaucracy: Empirical Evidence," THE BELL JOURNAL OF ECONOMICS AND MANAGEMENT SCIENCE, 7, 55-72.
McFadden, D. (1976) "Quantal Choice Analysis: A Survey," ANNALS OF ECONOMIC AND SOCIAL MEASUREMENT, 5, 363-390.
McFadden, D. (1976) "The Mathematical Theory of Demand Models," in P. Stopher and A. Meyburg (eds.), BEHAVIORAL TRAVEL-DEMAND MODELS, 305-314, D.C. Heath and Co.: Lexington, MA.
McFadden, D., A. Talvitie, & Associates (1977) "Demand Model Estimation and Validation," with A.P. Talvitie and Associates, URBAN TRAVEL DEMAND FORECASTING PROJECT, FINAL REPORT, VOLUME V, Institute of Transportation Studies, University of California, Berkeley.
McFadden, D. (1978) "Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments," in D. Hensher and P. Stopher (eds.), BEHAVIOURAL TRAVEL MODELLING, 279-318, Croom Helm London: London.
McFadden, D. (1978) “The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation," in EMERGING TRANSPORTATION PLANNING METHODS, U.S. Department of Transportation DOT-RSPA-DPB-50-78-2. Reprinted in T.H. Oum, et al. (eds.), TRANSPORT ECONOMICS: SELECTED READINGS, 51-80, Seoul Press: Seoul, 1995.

McFadden, D. (1978) Modelling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), SPATIAL INTERACTION THEORY AND PLANNING MODELS, 75-96, North Holland: Amsterdam. Reprinted in J. Quigley (ed.), THE ECONOMICS OF HOUSING, Vol. I, 531-552, Edward Elgar: London, 1997.

McFadden, D. (1981) "Econometric Models of Probabilistic Choice," in C.F. Manski and D. McFadden (eds.), STRUCTURAL ANALYSIS OF DISCRETE DATA WITH ECONOMETRIC APPLICATIONS, 198-272, MIT Press: Cambridge.

McFadden, D. (1984) "Econometric Analysis of Qualitative Response Models," in Z. Griliches and M. Intrilligator (eds.), HANDBOOK OF ECONOMETRICS,. II, 1396-1456, Elsevier: Amsterdam.

McFadden, D. (1986) "The Choice Theory Approach to Market Research," MARKETING SCIENCE, 5, 275-297.

McFadden, D. (1987) "Regression-Based Specification Tests for the Multinomial Logit Model," JOURNAL OF ECONOMETRICS, 34, 63-82.

McFadden, D. (1989) "Econometric Modeling of Locational Behavior," ANNALS OF OPERATIONS RESEARCH: FACILITY LOCATION ANALYSIS: THEORY AND APPLICATIONS, 18, 3-16.

McFadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," ECONOMETRICA, 57, 995-1026.

McFadden, D. et al (1991) "Advances in Computation, Statistical Methods, and Testing of Discrete Choice Models,"MARKETING LETTERS, 2, 215-229.

McFadden, D. (1998) "Measuring Willingness-to-Pay for Transportation Improvements" in T. Gärling, T. Laitila, and K. Westin (eds.) THEORETICAL FOUNDATIONS OF TRAVEL CHOICE MODELING, 339-364, Elsevier Science: Amsterdam.

McFadden, D. (2000) STATISTICAL TOOLS, manuscript.

McFadden, D. & F. Reid (1975) " Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models," with F. Reid, TRANSPORTATION RESEARCH RECORD: TRAVEL BEHAVIOR AND VALUES, 534, 24-37.

McFadden, D. , F. Reid, A. Talvitie, M. Johnson, and Associates (1979) "Overview and Summary: Urban Travel Demand Forecasting Project," with F. Reid, A. Talvitie, M. Johnson, and Associates, THE URBAN TRAVEL DEMAND FORECASTING PROJECT, FINAL REPORT, VOLUME I, Institute of Transportation Studies, University of California, Berkeley.

McFadden, D. & K. Train (1978) "The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models," TRANSPORTATION RESEARCH, 12, 349-353,.

McFadden, D., W. Tye, & K. Train (1978) "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," with W. Tye and K. Train, TRANSPORTATION RESEARCH RECORD: FORECASTING PASSENGER AND FREIGHT TRAVEL, 637, 39-46.

McFadden, D. & T. Morikawa (1986) "Discrete Response to Unobserved Variables for which there are Multiple Indicators," Dept. of Economics, MIT.

McFadden, D. & K. Train (1998) "Mixed MNL Models for Discrete Response," JOURNAL OF APPLIED ECONOMETRICS, forthcoming.

Meyer, J., and M. Straszheim (1970) "Transport Demand: The Basic Framework," TECHNIQUES OF TRANSPORTATION PLANNING, Brookings Institution: Washington, pp. 99-109; reprinted in T. Oum et al (eds) TRANSPORT ECONOMICS: SELECTED READINGS, Korea Research Foundation, 1995.

Morikawa, T. (1989) INCORPORATING STATED PREFERENCE DATA IN TRAVEL DEMAND ANALYSIS, Ph.D. Dissertation, Department of Civil Engineering, MIT.

Morikawa, T., M. Ben-Akiva, K. Yamada (1991) "Forecasting Intercity Rail Ridership Using Revealed Preference and Stated Preference Data," TRANSPORTATION RESEARCH RECORD, 1328, 30-35.

O'Donoghue, T. & M. Rabin (1999) "The Economics of Immediate Gratification," Dept. of Economics, Univ. of California, Berkeley

Papola, A. (2000) "Some Developments on the Cross-Nested Logit Model," Working Paper, Dipartimento di Ingegneria dei Tranporti, University of Naples..

Prelec, D. (1991) "Values and principles: Some limitations on traditional economic analysis," in A. Etzioni & P. Lawrence (eds) PERSPECTIVES ON SOCIOECONOMICS. London: M.E. Sharpe.

Quandt, R. (1970) THE DEMAND FOR TRAVEL: THEORY AND MEASUREMENT, Heath: Lexington.

Quarmby, D. (1967) "Choice of Travel Mode for the Journey to Work: Some Findings," Journal of Transportation Economics and Planning, 1, 273-314.

Revelt, D. & K. Train (1998) "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," REVIEW OF ECONOMICS AND STATISTICS, 80, 1-11.

Ruud, P. & D. McFadden (1994) "Estimation by Simulation," THE REVIEW OF ECONOMICS AND STATISTICS, 76, 591-608.

Small, K. (1987) "A Discrete Choice Model for Ordered Alternatives," ECONOMETRICA 55, 409-424.

Small, K. (1992) URBAN TRANSPORTATION ECONOMICS, Harwood: Chur.

Srinivasan, K., H. Mahmassani (2000) "Kernel Logit Method for the Longitudinal Analysis of Discrete Choice Data: Some Numerical Experiments," Working Paper, Department of Civil Engineering, University of Texas, Austin.

Stopher, P. & T. Lisco (1970) "Modeling Travel Demand: A Disaggregate Behavioral Approach, Issues and Applications," TRANSPORTATION RESEARCH FORUM PROCEEDINGS, 195-214.

Svenson, O. (1998) "The Perspective from Behavioral Decision Theory on Modeling Travel Choice," in T. Gärling, T. Laitila, and K. Westin (eds.) THEORETICAL FOUNDATIONS OF TRAVEL CHOICE MODELING, 141-172, Elsevier Science: Amsterdam.

Theil, H. (1969) “A Multinomial Extension of the Linear Logit Model,” INTERNATIONAL ECONOMIC REVIEW, 10, 251-259.
Thurstone, L. (1927) “A Law of Comparative Judgment,” PSYCHOLOGICAL REVIEW, 34, 273-286.
Train, K. (1986) QUALITATIVE CHOICE ANALYSIS, MIT Press: Cambridge.
Train, K. (1998) "Recreation Demand Models with Taste Differences over People," LAND ECONOMICS, 74, 230-239.
Train, K. (1999) “Halton Sequences for Mixed Logit,” Dept. of Economics, Univ. of California, Berkeley.
Train, K. and D. McFadden (1978) The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models," TRANSPORTATION RESEARCH, 12, 349-353.
Train, K.; D. McFadden; A. Goett, (1987) “Consumer Attitudes and Voluntary Rate Schedules for Public Utilities,” REVIEW-OF-ECONOMICS-AND-STATISTICS; 69, 383-91.
Warner, S. (1962) STOCHASTIC CHOICE OF MODE IN URBAN TRAVEL: A STUDY IN BINARY CHOICE, Northwestern University Press: Evanston.
Williams, H. (1977) “On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit,” ENVIRONMENT & PLANNING, A9, 285-344.
Wilson, A. (1971) “A Family of Spatial Interaction Models,” ENVIRONMENT AND PLANNING, 3.
Zachery, S. (1977) “Some Results on Choice Models,” Local Government Operations Research Unit Transportation Working Note 10.