ECONOMIC CHOICES

Prize Lecture, December 8, 2000

by

DANIEL L. MCFADDEN*

Department of Economics, University of California, Berkeley, CA 94720-3880, USA.

This Nobel lecture discusses the microeconometric analysis of choice behavior of consumers who face discrete economic alternatives. Before the 1960's, economists used consumer theory mostly as a logical tool, to explore conceptually the properties of alternative market organizations and economic policies. When the theory was applied empirically, it was to market-level or national-accounts-level data. In these applications, the theory was usually developed in terms of a *representative agent*, with market-level behavior given by the representative agent's behavior writ large. When observations deviated from those implied by the representative agent theory, these differences were swept into an additive disturbance and attributed to data measurement errors, rather than to unobserved factors within or across individual agents. In statistical language, traditional consumer theory placed structural restrictions on mean behavior, but the distribution of responses about their mean was not tied to the theory.

In the 1960's, rapidly increasing availability of survey data on individual behavior, and the advent of digital computers that could analyze these data, focused attention on the variations in demand across individuals. It became important to explain and model these variations as part of consumer theory, rather than as *ad hoc* disturbances. This was particularly obvious for discrete choices, such as transportation mode or occupation. The solution to this problem has led to the tools we have today for microeconometric analysis of choice behavior. I will first give a brief history of the development of this subject, and place my own contributions in context. After that, I will discuss in some detail more recent developments in the economic theory of choice, and modifications to this theory that are being forced by experimental evidence from cognitive psychology. I will close with a survey of statistical methods that have developed as part of the research program on economic choice behavior.

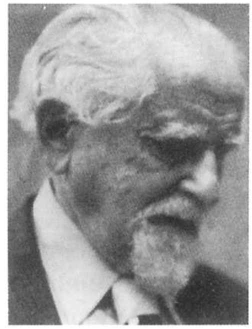
Science is a cooperative enterprise, and my work on choice behavior reflects not only my own ideas, but the results of exchange and collaboration with many other scholars.[1] First, of course, is my co-laureate James Heckman,

* Many of the author's publications cited in this paper are posted at http://elsa.berkeley.edu/~mcfadden.

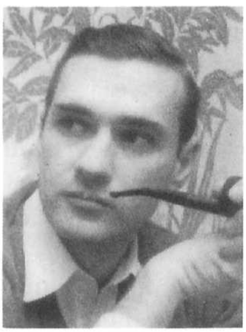
[1] Any accounting of credit for my contributions to economics has to include Leo Hurwicz, John Chipman, Marc Nerlove, and Hirofumi Uzawa, who attracted me to the field and taught me most of what I know.

Zvi Griliches

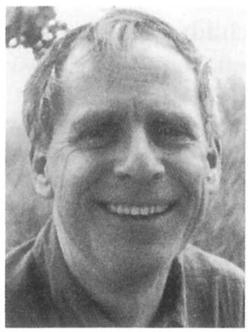
L. L. Thurstone



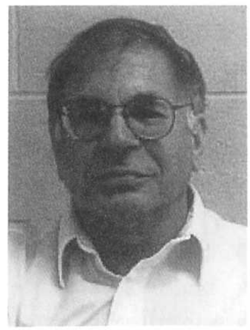
Jacob Marschak



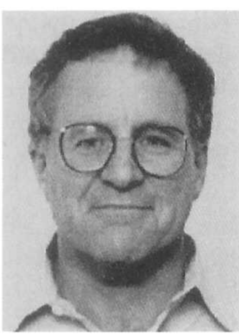
Duncan Luce



Amos Tversky



Danny Kahneman



Moshe Ben-Akiva

Charles Manski

Kenneth Train

Figure 1. Contributors to Discrete Choice Analysis.

who among his many contributions pioneered the important area of dynamic discrete choice analysis. Nine other individuals who played a major role in channeling microeconometrics and choice theory toward their modern forms, and had a particularly important influence on my own work, are Zvi Griliches, L.L. Thurstone, Jacob Marschak, Duncan Luce, Danny Kahneman, Amos Tversky, Moshe Ben-Akiva, Charles Manski, and Kenneth Train. A gallery of their photographs is shown in Figure 1. I wish particularly to cite Griliches, Marschak, and Tversky, robbed by death of their own chances to win Nobel prizes.

II. A BRIEF HISTORY

Classical economic theory postulates that consumers seek to maximize their self-interest, and that self-interest has broadly defined consistency properties across different decisions. At one level, the theory is virtually tautological, as in this description from a principles textbook by Frank Taussig (1912):

"An object can have no value unless it has utility. No one will give anything for an article unless it yield him satisfaction. Doubtless people are sometimes foolish, and buy things, as children do, to please a moment's fancy; but at least they think at the moment that there is a wish to be gratified."

The concept of rational consumer behavior was given a much more specific meaning in the perfection of the classical theory by John Hicks and Paul Samuelson, where self-interest is defined in terms of stable, innate preferences, and in Herb Simon's words, "The rational man of economics is a maximizer, who will settle for nothing less than the best."

Theorists considered heterogeneous preferences, but this complication was ignored in empirical studies of market demand that employed the representative consumer device. A consumer with preferences represented by a utility function $U(x)$ of a vector x of consumption levels of various goods would maximize this utility subject to a budget constraint $px \leq a$, where p is a vector of prices and a is income, at a demand function $x = d(a,p)$. This mapping was then assumed to hold at the market level with a disturbance ε added to account for discrepancies in observed data, $x = d(a,p) + \varepsilon$. The disturbance was interpreted as coming from measurement error in x, or possibly from consumer mistakes in optimization. Only representative demand $d(a,p)$ carried restrictions imposed by consumer theory.

The rapidly increasing availability of microeconomic data in the 1960's led econometricians to consider more carefully the specification of individual agent behavior. In 1957, Zvi Griliches pointed out that random elements appearing in the constraints or objectives of economic agents would produce disturbances in observed behavior whose properties depended on their source and whether they were known to the agents (Griliches , 1957; Mundlak, 1963; Griliches and Ringstad, 1970). I began working on these problems in 1962, in a study of production functions for electricity (McFadden, 1978a; Fuss, McFadden, and Mundlak, 1978).

In 1965, a Berkeley graduate student, Phoebe Cottingham, asked me for suggestions on how she might analyze her thesis data on freeway routing choices by the California Department of Highways. The problem was to devise a computationally tractable model of economic decision making that yielded choice probabilities $P_C(i)$ for the alternatives i in a finite feasible set C.

I was familiar with the work of psychologists on discrete choice behavior, and that seemed a promising place to start.

In a seminal paper on psychophysical discrimination, L.L. Thurstone (1927) introduced a *Law of Comparative Judgment* in which alternative i with true stimulus level V_i is perceived with a normal error as $V_i + \varepsilon_i$. The choice probability for a paired comparison then satisfied $P_{\{1,2\}}(1) = \Phi(V_1 - V_2)$, a form now called the binomial probit model. When the perceived stimuli $V_i + \varepsilon_i$ are interpreted as levels of satisfaction, or utility, this can be interpreted as a model for economic choice. Thurstone's work was introduced into economics by Jacob Marschak (1960), who explored the theoretical implications for choice probabilities of maximization of utilities that contained random elements. Marschak called this the *Random Utility Maximization* (RUM) model.

An influential study of choice behavior by R. Duncan Luce (1959) introduced an *Independence from Irrelevant Alternatives* (IIA) axiom that simplified experimental collection of choice data by allowing multinomial choice probabilities to be inferred from binomial choice experiments. The IIA axiom states that the ratio of choice probabilities for alternatives i and j is the same for every choice set C that includes both i and j; i.e., $P_C(i)/P_C(j) = P_{\{i,j\}}(i)/P_{\{i,j\}}(j)$.[2] Luce showed for positive probabilities that IIA implies *strict utilities* w_i such that $P_C(i) = w_i / \Sigma_{k\varepsilon C} w_k$. Marschak proved for a finite universe of objects that IIA implies RUM.

I proposed for Cottingham's research an econometric version of the Luce model in which the strict utilities were specified as functions of observed attributes of the alternative freeway routes,

$$(1) \qquad P_C(i) = \exp(V_i)\Sigma_{k\epsilon C}\exp(V_k).$$

In this formula, V_k was a *systematic utility* that I took to be a linear function of measured attributes of alternative k, such as construction cost, route length, and areas of parklands and open space taken, with coefficients that reflected the tastes of the decision-makers, and C was a finite set containing the feasible choice alternatives. I called this a *conditional logit model* since in the case of binomial choice it reduced to the logistic model used in biostatistics, and in the multinomial case it could be interpreted as the conditional distribution of demand given the feasible set of choice alternatives C. Today, (1) is more commonly called the *multinomial logit* (MNL) model, and I will use this more common terminology. I developed a computer program to estimate the MNL model by maximum likelihood, a non-trivial task in those days, and Cottingham completed her thesis before the program was working (Cottingham, 1966). However, I was eventually able to use the model to analyze her data (McFadden, 1968, 1976).

The characterization of alternatives in the MNL model in terms of their "hedonic" attributes was natural for this problem, and followed the psychometric tradition of describing alternatives in terms of physical stimuli. In em-

[2] The axiom can also be written as $P_C(i) = P_A(i) \cdot P_C(A)$ for $i \in A \subseteq C$, a variant that allows some alternatives to have a zero probability of being chosen.

pirical consumer theory, this was an early implementation of the hedonic formulation of the consumer problem developed by Griliches (1961) and Lancaster (1966).

As part of my development of the MNL model, I investigated further its RUM foundations. I showed that the Luce model was consistent with a RUM model with independent identically distributed additive disturbances if and only if these disturbances had a distribution called Extreme Value Type I. Earlier and independently, Tony Marley had established sufficiency (Luce and Suppes, 1965). Ket Richter and I also established a general necessary and sufficient condition for choice probabilities to be consistent with RUM, an *Axiom of Revealed Stochastic Preference* (ARSP): choice probabilities are RUM-consistent if and only if for any finite sequence of events (C_k, i_k), where C_k is a set of feasible alternatives and i_k is a choice, the sum of the choice probabilities does not exceed the maximum number of these events consistent with a single preference order (McFadden and Richter, 1970,1990).

Viewed as a statistical model for discrete response, the MNL model was a small and in retrospect obvious contribution to microeconometric analysis, although one that has turned out to have many applications. The reason my formulation of the MNL model has received more attention than others that were developed independently during the same decade seems to be the direct connection that I provided to consumer theory, linking unobserved preference heterogeneity to a fully consistent description of the distribution of demands (McFadden, 1974a).

I had an opportunity to develop additional applications of discrete choice analysis during a visit to M.I.T. in 1970. At that time, Peter Diamond and Robert Hall had developed a separable-utility, multi-stage budgeting, representative consumer model for the complex of consumer transportation decisions, including commute mode choice, and frequency, timing, and destination of shopping trips. They invited me to operationalize their model so that it could be estimated from data on individual trip-taking behavior. I did so using a nested version of the MNL model, with the nesting levels corresponding to the separable utility structure and with *inclusive values* carrying the impact of lower level decisions into higher levels in the same way that sub-budgets are carried through multi-stage budgeting problems (McFadden, 1974b; Domencich and McFadden, 1975). My treatment of inclusive values turned out to be approximately right, but a superior exact formula for inclusive values, utilizing what has come to be known as the *log sum formula*, was discovered by Ben-Akiva (1972).

Beginning in 1972, I organized a large research project at Berkeley, with support from the National Science Foundation, for the purpose of developing tools for transportation planning based on microeconometric analysis of individual travel decisions. Participants included Kenneth Train and Charles Manski. As a natural experiment to test and refine nested MNL models and other empirical RUM models, my research group studied the impact of BART, a new fixed-rail rapid transit system being built in the San Francisco Bay Area. We collected data on the travel behavior of a sample of individuals in 1972, prior to the introduction of BART, and estimated models that were then used to predict the behavior of the same individuals in 1975 after BART began operation. Table 1 summarizes results for the journey-to-work.

Table 1. Prediction Success Table, Journey-to-Work

(Pre-BART Model and Post-BART Choices)

Cell Counts	Predicted Choices				
Actual Choices	**Auto Alone**	**Carpool**	**Bus**	**BART**	**Total**
Auto Alone	255.1	79.1	28.5	15.2	378
Carpool	74.7	37.7	15.7	8.9	137
Bus	12.8	16.5	42.9	4.7	77
BART	9.8	11.1	6.9	11.2	39
Total	352.4	144.5	94.0	40.0	631
Predicted Share	55.8%	22.9%	14.9%	6.3%	
(Std. Error)	*(11.4%)*	*(10.7%)*	*(3.7%)*	*(2.5%)*	
Actual Share	59.9%	21.7%	12.2%	6.2%	

In this table, a MNL model estimated using the pre-BART commuter data was evaluated at the realized attributes of the alternatives, including the new BART alternative, that were available to each of the 631 subjects who were surveyed after BART began operation. The cell counts are the sums of the predicted probabilities for the sample individuals classified by their actual post-BART choice. The standard errors in the predicted shares are calculated taking into account the precision of model parameter estimates.

There were some systematic errors in our predictions. We over-estimated willingness to walk to BART, and under-estimated willingness to drive alone. In retrospect, the methods we used to assign an alternative-specific effect for the new BART mode, and to account for substitution between modes, were much inferior to the market research and modeling methods that are used today. However, our overall forecasts for BART were quite accurate, particularly in comparison to the official 1973 forecast, obtained from aggregate gravity models, that BART would carry 15 percent of commute trips. We were lucky to be so accurate, given the standard errors of our forecasts, but even discounting luck, our study provided strong evidence that disaggregate RUM-based models could out-perform conventional methods. Our procedures were also more sensitive to the operational policy decisions facing transportation planners. On the basis of our research, and other studies of the effectiveness of RUM-based travel demand analysis, these methods have been widely adopted for transportation planning around the world. Details of our research are found in (McFadden, Talvitie, *et al.*, 1977; McFadden, 1978b). The obvious similarities between the travel demand problem and applications such as education and occupation choices, demand for consumer goods, and location choices, have led to adoption of these methods in a variety of studies of choice behavior of both consumers and firms.

III. REFINEMENTS OF ECONOMIC CHOICE ANALYSIS

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 2.

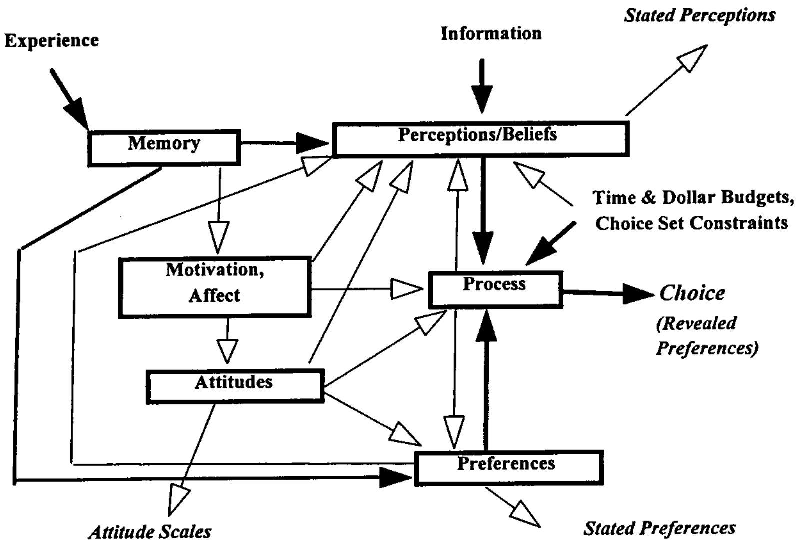


Figure 2. The Choice Process.

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 economists' *standard model* of the choice process, a theory of rational choice in which individuals collect information on alternatives, use the rules of probability to convert this information into perceived attributes, and then go through a cognitive process that 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; these I will discuss later. 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.

A. Fundamentals

The heart of the standard or rational model of economics is the idea that consumers seek to maximize innate, stable preferences whose domain is the vector of quantities and attributes of the commodities they consume. This holds 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., travel is an input to employment, and shopping activities are inputs to household production. An important feature of the theory is the *consumer sovereignty* property that preferences are predetermined in any choice situation, and do not depend on what alternatives are available. Succinctly, *desirability precedes availability.*

The standard model has a vaguely biological flavor. Preferences are determined from a genetically-coded *taste template.* The model allows experience to influence how preferences consistent with the template are expressed. However, most applications of the standard model leave out dependence on experience, and much of the power of this model lies in its ability to explain most patterns of economic behavior *without* having to account for experience or perceptions.

The original formulation of RUM as a behavioral hypothesis started from the standard model, with randomness attributed to unobserved heterogeneity in tastes, experience, and information on the attributes of alternatives. Parameterizing the utility function and the distribution of the random factors yielded parametric models for the choice probabilities, conditioned on observed attributes of alternatives and characteristics of the decision-maker. The MNL model is a tractable example. It is useful to review this derivation of the RUM explanation of choice 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 economic market data with experimental data on preferences, or who want to bring in cognitive and psychometric effects that are ignored in the standard model.

In the standard model, consumers have preferences over levels of consumption of goods and leisure. When goods have hedonic attributes, preferences are defined to incorporate the consumer's subjective perceptions of these attributes. The expressed preferences of the consumer are functions of their taste template, experience, and personal characteristics, including both observed and unobserved components. Mild regularity conditions allow us to represent preferences by a continuous real-valued utility function of the characteristics of the consumer, and consumption levels and attributes of goods. Consumers are heterogeneous in unobserved characteristics such as their taste templates and the mechanisms they use to form perceptions. I will assume that the unobserved characteristics vary continuously with the observed characteristics of a consumer. For example, the tastes and perceptions of an individual change smoothly with age as long as there are no major shifts in observed characteristics. Technically, this is an assumption that unobserved characteristics are a *continuous random field* indexed by the observed characteristics. An implication of this assumption is that the conditional distribution of the unobserved characteristics will depend continuously on the observed characteristics. This assumption is not very restrictive, and can essentially be made true by construction.

One important restriction that consumer sovereignty places on the conditional distribution of unobserved consumer characteristics is that it cannot depend on *current* economic variables such as non-wage income, the wage rate, and goods prices, which determine feasibility through the consumer's budget, but are excluded from influencing tastes. The conditional distribution can however depend on the individual's *history* of economic status and choices, through the operation of experience on the expression of preferences. Under mild regularity conditions, the random field of unobserved consumer characteristics can be written as a continuous transformation of a *uniform* continuous random field; this is an extension of an elementary result from probability theory that a univariate random variable Y with distribution F can be written almost surely as $Y = F^{-1}(\upsilon,)$ with υ, a uniform (0,1) random variable. This transformation can then be absorbed into the definition of the utility function, so that the dependence of the utility function on unobserved consumer characteristics can be represented canonically as a continuous function of a uniformly distributed random vector.

I consider discrete choice from feasible sets containing finite numbers of mutually exclusive and exhaustive alternatives that are characterized by their observed attributes, with other aspects of consumer behavior taking place in the background. Suppose for the moment that the consumer is assigned a specific discrete alternative. Given this alternative, non-wage income net of the cost of the alternative, the wage rate, and goods prices, the consumer will choose leisure and consumption levels of remaining goods to maximize utility subject to budget and time constraints. The level of utility attained is then a function of the attributes of the discrete alternative, observed consumer characteristics, a uniformly distributed random vector characterizing unobserved consumer characteristics, and the economic variables that determine the budget constraint: net non-wage income, the wage rate, and goods prices. The theory of optimization implies that this is a classical indirect utility function, with the properties that it has a closed graph and is quasi-convex and homogeneous of degree zero in the economic variables, and increasing in net non-wage income. Under fairly mild conditions, it is possible to require that the indirect utility function be convex, rather than quasi-convex, in the economic variables. The last step in applying the standard model to discrete choice is to require the consumer's choice among the feasible alternatives to maximize conditional indirect utility.

The functional form of the canonical indirect utility function will depend on the structure of preferences, including the trade-off between goods and leisure as non-wage income or the wage rate change, the role of household production in determining how goods combine to satisfy needs, and separability properties of preferences. The original 1970 formulation of the RUM model for travel demand applications fit into this framework, in some variant of the form

$$(2) \qquad U = V + \eta \text{ and } V = [\alpha\cdot(a\text{-}c)/w - \beta\cdot t]\cdot w^{\theta} + z(x,s)\cdot\gamma.$$

In this formula, a is non-wage income, c is the cost of the alternative, w is the wage rate, with (a,c,w) all expressed in real terms with other goods prices implicit, t is the time required by the alternative, x is a vector of other observed attributes of the alternative, s is a vector of observed characteristics of the consumer, and $z(x,s)$ is a vector of pre- specified functions of the arguments. The (α,β,γ) are parameters, and θ determines the elasticity of the demand for leisure and is commonly assumed to be either zero or one, but can be a parameter in (0,1) corresponding to a Stone-Geary specification for systematic utility (McFadden and Train, 1978). The η is an additive disturbance summarizing the effects of unobserved consumer characteristics. When η = -log(-log(ε)) and the ε are uniformly distributed and independent across alternatives, the disturbances are independently identically extreme value distributed and produce a MNL model (1) in which the systematic utility has the form (2) for each $k \in C$.

A natural question to ask in retrospect is how restrictive this specification is, and to what degree it can be modified to accommodate more general RUM-consistent behavior. The answer is that both the linear dependence of systematic utility on economic variables and the distributional assumption yielding the IIA property are quite special. While the model works well as an an empirical approximation in surprisingly many applications, it implies a uniform pattern of substitution between alternatives that may not be behaviorally plausible. A number of more flexible and more or less tractable families of models have been developed with more general dependence on explanatory variables and/or distributions of unobservables that permit more general patterns of substitution between alternatives.

B. Models for RUM-Consistent Choice Probabilities

The MNL model has proven to have wide empirical applicability, but as a theoretical model of choice behavior its IIA property is unsatisfactorily restrictive. Examples due to Chipman (1960) and Debreu (1960), later elaborated as the "red-bus, blue-bus" problem in transportation applicacions, show that we can sometimes expect this model to fail. *Nested MNL* models, *generalized extreme value* (GEV) models, and *multinomial probit* (MNP) models have been developed to relax the restrictive properties of the simple MNL model. These are often very useful, but remain restrictive in the sense that tractable versions fall short of being able to represent all RUM-consistent behavior. One family of RUM-consistent discrete choice models that is very flexible is the random parameters or mixed multinomial logit (MMNL) model.

GEV models were introduced and their RUM consistency established in McFadden (1978b). Define a *GEV generating function* $H(w_1,\ldots,w_J)$ to be a non-negative linear homogeneous function of $w \geq 0$ with the property that H goes to $+\infty$ when any argument goes to $+\infty$, and with continuous mixed partial derivatives that alternate in sign, with non-negative odd mixed derivatives. Then $F(\eta_1,\ldots,\eta_J) = \exp(-H(e^{-\eta_1},\ldots,e^{-\eta_J}))$ is a joint distribution function whose one-dimensional marginals are extreme value distributions. Consider a RUM model $u_i = V_i + \eta_i$ for a set of alternatives $C = \{1,\ldots,J\}$, where the η's have this distribution. Then $E\ max_i\ u_i = \log(H(e^{V_1},\ldots,e^{V_J})) + \zeta$, where $\zeta = 0.57721$ ... is Euler's constant. The RUM choice probabilities are given by the derivatives of this expectation, with the closed form

(3) $$P_C(i) = e^{V_i} \cdot H_J(e^{V_1},...,e^{V_J}) / H(e^{V_1},...,e^{V_J}).$$

One example of a GEV generating function is the linear function $H = w_1 + ... + w_J$; this yields the MNL model. More complex GEV models are obtained by repeated application of the following result: If sets A,B satisfy $A \cup B = C$, and w_A, w_B, and w_C are the corresponding subvectors of $(w_1,...,w_J)$, if $H^A(w_A)$ and $H^B(w_B)$ are GEV generating functions in w_A and w_B, respectively, and if $s \geq 1$, then $H^C(w_C) = H^A(w_A{}^s)^{1/s} + H^B(w_B)$ is a GEV generating function in w_C. The parameter $1/s$ is called an *inclusive value coefficient. Nested MNL* models are defined by applying this recursion repeatedly to non-overlapping sets A and B, and the argument shows they are RUM-consistent.

Mixtures of RUM-consistent choice models are again RUM-consistent. For example, if $H(w_1,...,w_J,\alpha)$ is a family of GEV generating functions indexed by parameters α that determine nesting structure, weights, and inclusive values, and one has a distribution over α that does not depend on economic variables, then the RUM model $u_i = V_i + \eta_i$ with $F(\eta_1,...,\eta_J) = E_\alpha \exp(-H(e^{-\eta_1},...,e^{-\eta_J};\alpha))$ has $E \max_i u_i = E_\alpha log(H(e^{V_1},...,e^{V_J};\alpha)) + \zeta$ and choice probabilities satisfying $P_C(i) = \partial E\ max_i\ u_i / \partial V_i = E_\alpha e^{V_i} \cdot H_i(e^{V_1},...,e^{V_J},\alpha)/H(e^{V_1},...,e^{V_J},\alpha)$. Useful specializations of the GEV family can be found in McFadden (1981); Small (1987); Bhat (1998).

A different approach that established the RUM-consistency of an important family of nested MNL models was taken by Williams, (1977); Daly and Zachary (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. Second, they effectively showed that in the family of RUM models with an additive linear non-wage 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. 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. Proofs of these results are given in McFadden (1981); McFadden and Train (2000).

Generalized extreme value families of choice models avoid some IIA restrictions, but cannot represent all RUM-consistent behavior. The MNP model, obtained from a RUM model with additive normal disturbances that have a general covariance structure is quite flexible, but its choice probabilities must usually be written in open form as multivariate integrals that require numerical integration. Special restrictions such as factor-analytic covariance structures are needed to make these models tractable (McFadden, 1981, 1984). However, simulation-based estimation methods, discussed later, have improved our ability to implement fairly general forms of these models in applications.

Recently, McFadden and Train, (2000) have established a somewhat surprising and convenient *mixed MNL* (MMNL) approximate representation of any regular RUM-consistent choice probabilities. Start from the canonical representation of the standard model described earlier. Make the fairly mild assumption that the class of all feasible sets is compact. Perturb the canonical indirect utility functions by adding independent Extreme Value Type I disturbances, scaled so that the probability is very small that the original and perturbed indirect utility functions order alternatives differently. Further, approximate the canonical indirect utility uniformly by a Bernstein-Weierstrauss polynomial in the observed arguments and the uniformly distributed vector of unobserved characteristics.[3] This can again be done so that the probability of the approximation changing the preference order is very small. Condition on the uniform random vector that enters the utility function, and then integrate this vector out to obtain the MMNL model,

$$P_C(i) = \int_0^1 \ldots \int_0^1 \frac{e^{Z_i \cdot \alpha(\varepsilon)}}{\sum_{j \in C} e^{Z_j \cdot \alpha(\varepsilon)}} \cdot d\varepsilon \quad . \tag{4}$$

In this formula, $\boldsymbol{\alpha}(\varepsilon)$ is a vector of polynomial functions of the uniform random vector $\boldsymbol{\varepsilon}$, and the $\mathbf{Z}_j$ are vectors of polynomial functions of observed characteristics of the consumer and observed attributes of alternative j. It is immediate from its derivation that every MMNL model of the form (4) is RUM-consistent, provided the functions $\mathbf{Z}_j \cdot \alpha(\varepsilon)$ are indirect utility functions for each ε. The model (4) 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 systematic utility of an alternative by introducing a series approximation in the observed attributes of the alternative, interacted with observed characteristics of the decision-maker. In principle, the approximation errors in this formulation can be bounded and the order of the polynomial required to achieve a desired level of accuracy can be determined in advance. However, the quantities this calculation requires are often unavailable in applications, and it is better to use an adaptive or cross-validation method to determine a stopping point for the approximation. The shape restrictions required on $\mathbf{Z} \cdot \alpha(\varepsilon)$ are most easily imposed component-by-component, with sign restrictions on the corresponding components of $\boldsymbol{\alpha}$. Theoretically, it is possible to select a basis so that this can be done without losing the uniform approximation property, but this has been done constructively only for one and two dimensions (Anastassiou and Yu, 1992; Dechevsky and Penev, 1997). Alternately, one can proceed without imposing the shape restrictions, and test for them in the range of the observations (Brown and Matzkin, 1998).

One can approximate the distribution of the α coefficients in (4) by a distribution concentrated on a finite set of points, with the probability weights at

[3] Other Hamel bases for the approximation can also be used, and may have advantages in terms of parsimony and the imposition of shape restrictions.

these points treated as parameters. This is called a *latent class* model. It is possible to use latent class models to obtain non-parametric estimates of any family of RUM-consistent choice probabilities by the method of sieves. The latent class model is a *single hidden-layer feedforward neural network* (with MNL activation functions), and the asymptotic approximation theory that has been developed for neural networks can be applied to establish convergence rates and stopping rules (White, 1989,1992; Cheng and Titterington, 1994; Chen and White, 1999; Ai and Chen, 1999). It is possible to develop other RUM-consistent approximations to families of choice probabilities that are useful in some applications (Dagsvik, 1994).

Summarizing, I have outlined a result which says that any well-behaved RUM model can be approximated by a MMNL model, or alternately by a latent class model, provided the transformations of observed variables and the random distributions that enter these forms are sufficiently flexible. The MMNL model was introduced by Cardell and Dunbar (1980). With the development of convenient simulation methods for estimation (Revelt and Train, 1998), it has become widely used.

To illustrate application of the MMNL model, I will describe a study of trout fishing destination choice conducted as part of an assessment of damage that copper mining caused to recreational fishing in the Clark Fork River Basin in Montana. A sample of 962 fishing trips to 59 sites on Montana rivers, made by 238 anglers, was collected in a household survey conducted by Bill Desvousges and associates at Triangle Economic Research. The variables in the study are described in Table 2.

These data have been used by Train (1998) to estimate MMNL models of the form (4) for fishing site choice. This study assumes an indirect utility model $U = \alpha(a-c) - \beta wt + z(x,s)\gamma$, where the notation is the same as in (2), and the parameters (α,β,γ) vary randomly over the population, with a specification that fixes the ratio β/α and for the estimates described in Table 3 takes α and γ to have independently distributed components that are either normal

Table 2. Explanatory Variables for Trout Fishing Destination Choice in Montana

Variable	Description	Mean	Std. Dev.
Trout Stock	Hundreds of fish per 1000' of stream	1.773	1.468
Trip Cost	Travel cost to the site, including the variable cost of driving and the value of time spent driving (calculated at 1/3 the angler's wage, or $\beta = \alpha/3$)	\$89.22	\$35.24
Access	Number of State designated access areas per USGS block	0.172	0.305
Aesthetics	Rating 0 to 3 (Montana River Information System)	1.386	0.86
Campgrounds	Number of campgrounds per USGS block	0.195	0.198
Major	Major fishing area (Angler's Guide to Montana)	0.559	0.501
Restricted	Number of restricted species at the site (e.g., mandated catch/release) during some of year	0.339	0.902
Logsize	Log of number of USGS blocks that contain the site	2.649	0.684

or log normal. The table gives percentiles of the estimated parameter distributions. Notable in this model is the spread in the distribution of tastes for number of trout, which determines catch rates, and the division of anglers between positive and negative tastes for campgrounds and number of access points, which provide convenience but also produce crowding. The elasticity is the percentage increase in the probability for a site resulting from a one percent increase in the explanatory variable for that alternative, calculated at sample average values for the variables and the probabilities.

Table 3. MMNL Model of Fishing Site Choice with Independent Random Parameters

Variable	Parameter Distribution	Distribution of Coefficient			Proportion Positive	Elasticity (at Median Coefficient)
		10 Pctle	Median	90 Pctle		
Trout Stock	Log Normal	0.015	0.056	0.207*	1.0	0.098
	Std. Error	*0.014*	*0.034*	*0.070*		
Trip cost	Log Normal	-0.253*	-0.091*	-0.032*	0.0	-7.945
	Std. Error	*0.030*	*0.006*	*0.004*		
Access	Normal	-3.369*	-0.950*	1.470*	0.31	-0.161
	Std. Error	*0.715*	*0.361*	*0.392*		
Aesthetics	Log Normal	0.152*	0.452*	1.342*	1.0	0.616
	Std. Error	*0.060*	*0.103*	*0.159*		
Camp-grounds	Normal	-2.005*	0.116	2.237*	0.53	0.022
	Std. Error	*0.693*	*0.323*	*0.591*		
Major	Normal	-1.795*	1.018*	3.831*	0.68	0.56
	Std. Error	*0.401*	*0.289*	*0.642*		
Restricted	Normal	-1.651*	-0.499*	0.653*	0.29	-0.166
	Std. Error	*0.305*	*0.131*	*0.171*		
Logsize	Fixed	0.9835*	0.9835*	0.9835*	1.0	
	Std. Error	*0.108*	*0.108*	*0.108*		0.967

* = significant, 1% level

C. Estimating Willingness-to-Pay in Discrete Choice Models

Applications of discrete choice models to economic policy problems often call for estimation of Willingness-to-Pay (WTP) for policy changes. For example, the Montana trout fishing study sought to determine WTP for the increase in fish stocks that would come from restoration of natural riparian conditions. For the MMNL model in Table 3 which is independent of non-wage income, mean WTP has a convenient exact *expected log sum* form in the systematic utilities before (V_i') and after (V_i'') the change,

$$\text{WTP} = E_{\alpha,\beta,\gamma} \frac{1}{\alpha} \log\left(\frac{\exp(V_1'') + \ldots + \exp(V_J'')}{\exp(V_1') + \ldots + \exp(V_J')} \right) . \tag{5}$$

This is a case where Hicksian and Marshallian measures of consumer surplus coincide, and also where preferences can be aggregated into representative "community" preferences (Chipman and Moore, 1990; McFadden, 1999b).

When the indirect utility function is not linear and additive in non-wage income, computation of exact Hicksian compensating variation is much more burdensome. McFadden gives bounds that will sometimes suffice for policy analysis, and develops monte carlo markov chain methods for numerical calculation of exact WTP. Recently, Karlstrom (2000) has developed numerical methods that simplify these calculations.

D. Dynamic Models

A major opening up of the study of economic choice behavior occurs when one turns to data on repeated choices by the same individuals, and the dynamics of discrete choice. It is in this panel framework that the operation of experience on the evolution of perceptions and tastes, postulated in Figure 2, can be examined empirically. Repeated decisions also allow one to study RUM theory as an intra-consumer as well as an inter-consumer model of taste variation, providing a link to psychological models of decision-making. Analysis of the dynamics of discrete choice has been pioneered by Heckman (1981a,b), who recognized the critical roles of initial values and recursive structure in well-specified dynamic models and is responsible for the fundamental development of appropriate econometric methods. Dynamic models have important applications to issues of labor supply and job status, and also to a variety of subjects ranging from the timing of purchases of new goods to life-cycle decisions like retirement.

An important element in analysis of the dynamics of discrete choice has been the integration of expectations into choice decisions, through dynamic optimization (Rust, 1994) and through interactions between agents (Laffont and Vuong, 1996). The last topic is related to a more general issue in discrete choice analysis. In many microeconomic data sets, the explanatory variables behind an economic choice can be treated as predetermined, because the feedback from decisions of individual consumers to market-level economic variables is weak. However, in dynamic models where current unobservables are not necessarily independent of past history, or in "thin" markets where there is strategic interaction between agents, feedbacks become strong enough so that it is necessary to deal with endogeneity in explanatory variables.

E. Discrete/Continuous Choice

Discrete and continuous components of economic decisions are fully integrated in economic choice theory, through common preferences and time and budget constraints. However, this integration has rarely been carried through in empirical study of consumer behavior. Dubin and McFadden (1984) develop a consistent model of discrete and continuous decisions for application to choice and use of consumer products, but the cost of computational tractability is a highly restrictive parameterization. Further development of this topic, perhaps using semiparametric estimation to relax model restrictions, is needed.

IV. THE PSYCHOLOGY OF CHOICE BEHAVIOR

In psychological theories of the choice process, the individual is less organized, and more adaptive and imitative, than in the economists' standard model. Psychological descriptions of decision-making are both colorful and intuitive. *Attitudes* play a major role in determining how consumers define the decision-making task. In the words of Danny Kahneman, "Economists have preferences; psychologists have attitudes." *Affect* and *motivation* are key determinants of attitudes; and also influence the *perceptions* that feed into the choice process; see the light arrows in Figure 2. In these theories, the economists' calculus of utility assessment and maximization is reduced to one of many factors in the decision-making environment, with an influence that is often overridden by context effects, emotion, and errors in perception and judgment; see (Svenson, 1979; Garling, 1992; Lowenstein, 1996). Experimental evidence and self-reported decision protocols support the view that heuristic rules are the proximate drivers of most human behavior. The psychologist 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."

Human behavior may be governed by rules, but it is possible that these rules simply encode preferences. The evolutionary and behavioral arguments used to explain the reinforcement of self-protective rules systems also suggest that selection will favor rules systems that consistently advance self-interest. Many psychologists argue that behavior is far too sensitive to context and affect to be usefully related to stable preferences. However, if there are underlying preferences, then even if the link from preferences to rules is quite noisy it may be possible to recover these preferences and use them to correctly evaluate economic policies, at least as an approximation that is good enough for government policy work.

The existence of underlying preferences is a vital scientific question for economists. If the answer is affirmative, then the evidence on decision-making from cognitive psychology implies only that economists must look through the smoke-screen of rules to discern the deeper preferences that are needed to value economic policies. This is a difficult task, but not an impossible one. If the answer is negative, then economists need to seek a foundation for policy analysis than does not require that the concept of "greatest good for the greatest number" be meaningful. I am guardedly optimistic that the question has an affirmative answer. The first reason is that many behavioral deviations from the economists' standard model are explained by perceptual illusions and information processing errors, rather than a more fundamental breakdown in the definition of self-interest. The second is that many of the rules we do use are essentially defensive, protecting us from bad choices. To illustrate, consider the simplified road map of the wine-producing region around Bordeaux shown in Figure 3.

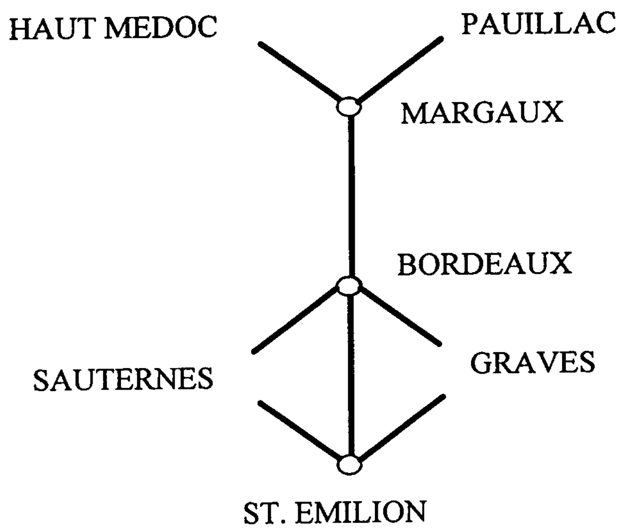


Figure 3. Roads in the Wine-Producing Region near Bordeaux.

Bordeaux appears to be closer to St. Emilion than to Margaux. However, you will immediately recognize that this is a version of the classical Muller-Lyer optical illusion in which the distances are actually the same. Even after you are reminded of this, St. Emilion looks closer. Could this illusion affect behavior? It may be significant that Figure 3 was adapted from a brochure published by the commune of St. Emilion. And in fact St. Emilion is more crowded than Margaux, perhaps as a result of enophiles' illusions. However, I doubt that this is due to mass misreading of maps by travelers to Bordeaux. We learn to be suspicious of our perceptions. We may see things cock-eyed, but we adopt conservative behavioral strategies, such as measuring map distances, that prevent us from deviating too far from our self-interest.

In light of this example, how should a scientist go about predicting travel decisions of map-readers? One place to start is the library of optical illusions. These certainly help to reveal the cognitive processes involved in vision.

However, it is very difficult to synthesize this library into a forecasting system that is broadly predictive. Another starting point is a crude "you see what a camera sees" model of vision. We know from the very existence of optical illusions that this model is not universally true. Despite this, the crude model is broadly predictive, and even more so if it is relaxed to accommodate some systematic illusions. I consider this a good analogy for economists deciding how to predict economic choice behavior. Until the day comes when brain science understands how the cognitive mechanisms operate in Figure 2 for a broad spectrum of economic decisions, I suspect that the standard model, enhanced to account for the most systematic perceptual illusions, will prove to be the best platform for evaluating most economic policies .

A. Cognitive Illusions

The preceding discussion has treated the psychological view of decision-making as a theoretical alternative to the standard model, but there is now also substantial evidence that in a laboratory setting individuals will sometimes make decisions that deviate strikingly and systematically from the predictions of the standard model. The experimental results of Danny Kahneman and Amos Tversky (e.g., Tversky and Kahneman, 1974, 1981; Kahneman and Tversky, 1979, 1984) have been particularly influential in forcing economists to rethink the standard model. Table 4, adapted from McFadden (1999a), lists some of the (overlapping) cognitive phenomena identified by cognitive psychologists and behavioral economists that appear to influence behavior.

One important cognitive phenomenon is *anchoring*, in which responses are pulled toward numerical prompts, even when they are uninformative (Tversky and Kahneman, 1974). A psychological explanation for anchoring is that a prompt creates in the subject's mind, at least temporarily, the possibili-

Table 4. Cognitive Effects in Decision-Making

Effect	Description
Anchoring	Responses are influenced by cues contained in the question
Availability	Responses rely too heavily on readily retrieved information, and too little on background information
Context	Previous questions and interviewer interaction color perception
Framing/Reference Point	Question format changes saliency of different aspects of the cognitive task
Focal	Quantitative information is stored and/or reported categorically
Primacy/Recency	Initial and recently experienced events are the most salient
Projection	Responses are consonant with the self-image the subject wishes to project
Prospect	The likelihoods of low probability events are misjudged, and treated either as too likely or as zero
Regression	Causality and permanence are attached to past fluctuations, and regression to the mean is underestimated
Representativeness	High conditional probabilities induce overestimates of unconditional probabilities
Rule-Driven	Motivation and self-control induce strategic responses
Saliency	The most salient aspects of the question are overemphasized
Status Quo	Current status and history are privileged
Superstition	Elaborate causal structures are attached to coincidences
Temporal	Temporally inconsistent time discounting

ty that the uncertain quantity could be either above or below the prompt. This could result from classical psychophysical discrimination errors, or from a cognitive process in which the subject treats the question as a problem-solving task and seeks an appropriate framework for "constructing" a correct solution. Evidence suggests that individuals are poor natural statisticians, placing too much weight on readily available information and exemplars, and too little on background information that is more difficult to retrieve. Education trains individuals to use problem-solving protocols in which responses to questions are based not only on substantive knowledge, but also on contextual cues as to what a correct response might be. Consequently, it is no surprise if subjects apply these protocols and use numerical prompts in forming responses.

B. Bias in Reported Consumption

I will describe two experiments that show anchoring is at least a problem for measurement in economic surveys. The first, taken from Hurd, Merrill, and McFadden (1997), is concerned with response bias when subjects are asked to report on economic quantities they may not know with certainty. These authors conducted an experiment in the AHEAD panel, a large study of the elderly in the United States. Subjects were asked about their monthly consumption, using an unfolding brackets format that asked for yes/no responses to a series of numerical prompts. The pattern of prompts given to each subject was selected by experimental design. For the range of initial prompts used in the experiment, from $500 per month to $5000 per month, this led the implied median consumption levels to vary from $895 per month to $1455 per month; see Figure 4. More detailed information on the experimental results is given in Table 5. The distributions of responses for the different treatment groups show convincingly that the anchoring phenomenon can introduce response bias that if unrecognized might seriously distort economic policy analysis.

C. Bias in Stated Willingness-to-Pay

The second study, by Green, Jakowitz, Kahneman, and McFadden (1998), asks subjects recruited from visitors to a science museum to state their willingness to pay to save off-shore seabirds from small oil spills . Subjects were assigned randomly to control and treatment groups. Both groups were given the following preamble:

There is a population of several million seabirds living off the Pacific coast, from San Diego to Seattle. The birds spend most of their time many miles away from shore and few people see them. It is estimated that small oil spills kill more than 50,000 seabirds per year, far from shore. Scientists have discussed methods to prevent seabird deaths from oil, but the solutions are expensive and extra funds will be required to implement them. It is usually not possible to identify the tankers that cause small spills and to force the companies to pay. Until this situation changes, public money would have to be spent each year to save the birds. We are interested in the value your household would place on saving about 50,000 seabirds each year from the effects of offshore oil spills.

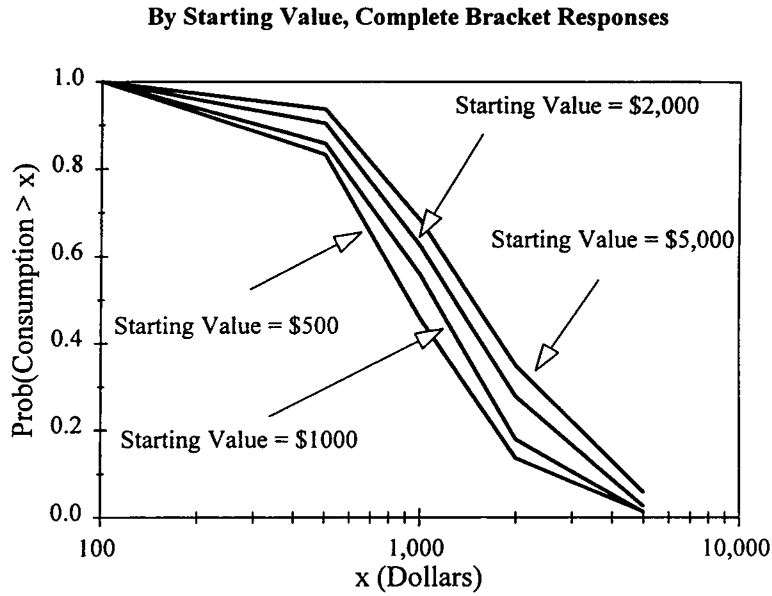


Figure 4. Monthly Consumption.
Reprinted from McFadden, Daniel, "Consumption and Savings Balances of the Elderly," with Michael D. Hurd, Harish Chand, Li Gan, Angela Merrill, and Michael Roberts, in – Frontiers in the Economics of Aging –, David A. Wise, editor, Chicago University Press, 1998.

The control group was then given this open-ended question:

If you could be sure that 50,000 seabirds would be saved each year, what is the MOST your household would pay in extra federal or state taxes per year to support an operation to save the seabirds? The operation will stop when ways are found to prevent oil spills, or to identify the tankers that cause them and make their owners pay for the operation.

$ __________ per year

The treatment groups were given the referendum question:
If you could be sure that 50,000 seabirds would be saved each year, would you agree to pay {$5} in extra federal or state taxes per year to support an operation to save the seabirds? The operation will stop when ways are found to prevent oil spills, or to identify the tankers that cause them and make their owners pay for the operation.

Yes No

This question was then followed up by an open-ended question

What is the MOST that you would be willing to pay? $ __________ per year

The numerical prompt of $5 in the referendum question was varied across several levels set by experimental design, with the treatments selected to correspond to specified quantiles of the control group's distribution of responses. If subjects conform to the economists' standard model, their preferences are innate and will not be anchored to the numerical prompts contained in the referendum questions. In fact, the response patterns suggest the prompt

Table 5. Consumption: Sample Sizes, Medians, and Means

Treatment	Starting Gate Amount	Sample Size	Number with Open-Ended Response	Percentage of Bracket Responses Completed	Medians: Nonparametric (b)	Medians: Std. Error (c)	Medians: Parametric (d)	Medians: Std. Error (e)	Means: Nonparametric (f)	Means: Std. Error (g)	Means: Parametric (h)	Means: Std. Error (i)
1	2000 (a)	739	492	53.8%	1061	87	1128	72	1732	108	1513	88
2	500 (a)	689	422	51.3%	861	53	864	53	1261	87	1139	63
3	1000	627	0	92.8%	1146	39	1104	37	1508	49	1365	40
4	5000	782	0	94.0%	1455	56	1486	52	2161	65	1979	62
5	500	707	0	92.9%	895	31	934	31	1311	45	1180	35
6	2000	594	0	94.1%	1415	53	1392	51	1946	61	1764	57
7	1000 (a)	717	464	47.0%	897	62	967	69	1466	98	1352	89
2&5	500	1396	422	81.5%	886	26	915	27	1298	40	1170	31
3&7	1000	1344	464	79.7%	1090	36	1066	33	1497	44	1364	38
1&6	2000	1333	492	82.3%	1326	46	1310	42	1884	53	1695	49
OE First (1,2,7)		2145	1378	50.7%	931	35	980	37	1485	57	1331	46
Forced (3,5,6)		1928	0	93.3%	1129	25	1167	25	1572	30	1523	29
Pooled (1,2,3,5,6,7)		4073	1378	81.2%	1077	22	911	18	1358	31	1237	22
Open-Ended Responses			1378		1000	9			1253	64		
Overall		4855	1378	84.0%	1163	21	1170	19	1696	26	1534	22

NOTES:

(a) Subjects were first asked for an open-ended response, with unfolding brackets if there was no response to the open-ended question.

(b) Exponential of linearly interpolated CCDF of log consumption, with the CCDF estimated using a "saturated" multinomial model for all respondents.

(c) The standard error is estimated by (median)x(a-b)/(2x(prob. of bracket)x(root N)), where (b,a) is the log consumption bracket containing the estimator. This estimator assumes that log consumption is uniformly distributed within the bracket containing the median.

(d) Exponential of the mean of a log normal distribution fitted by MLE to bracket frequencies of log consumption.

(e) The standard error is estimated by (median)x(SD)xroot(pi/2xN)., where SD is the estimated standard error of log consumption.

(f) Sum of bracket midpoints times estimated bracket probabilities.

(g) Standard Error is estimated by square root of (sum of squared bracket midpoints times bracket probabilities minus median squared)/N.

(h) Exponential of (mean)+0.5x(sigma)^2, where mean and sigma are estimates of the mean and standard deviation of log consumption.

(i) The standard error is estimated by (mean)x(SD)xroot(1+0.5x(SD)^2)/root(N), where SD is the estimated standard error of log consumption.

Reprinted from McFadden, Daniel, "Consumption and Savings Balances of the Elderly," with Michael D. Hurd, Harish Chand, Li Gan, Angela Merrill, and Michael Roberts, in –Frontiers in the Economics of Aging –, David A. Wise, editor, Chicago University Press, 1998.

creates an "anchor" for the construction of preferences. Figure 5 shows the differences in response frequencies for the control and treatment groups; details are given in Table 6. The anchoring effects are large, and statistically significant. Comparing open-ended and follow-up responses, the median WTP in the control group is \$25, and median WTP in the treatment group varied from \$10 when the prompt was \$5 to \$50 when the prompt was \$400. The referendum responses also show an anchoring effect, with higher pluralities for "yes" at higher prompts than in the control group. These produce a non-parametric estimate of \$167 for mean WTP in the treatment group, compared with a mean of \$64 in the control group, again statistically significant. Put another way, the effect of a one dollar increase in the prompt is to increase mean response by 28 cents. This experiment also showed that anchoring in response to the WTP question paralleled anchoring in responses to objective estimation questions, such as the height of the tallest redwood tree in California.

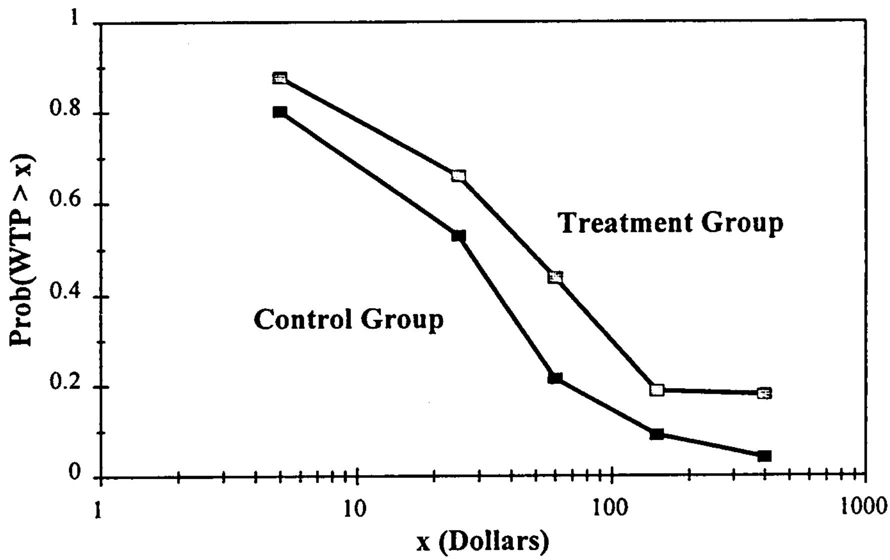


Figure 5. WTP for Seabirds.
Adapted from "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," D. McFadden with D. Green, K. Jacowitz, and D. Kahneman, Resource and Energy Economics, Vol. 20, 85–116, 1998. Copyright 1998. Reprinted with permission from Elsevier Science.

The Green *et al.* experiment was hypothetical, and subjects were aware that their responses would have no direct monetary consequences. A natural question for economists to ask is whether such deviations from the standard model continue to appear in market choices where real decisions involve real money. The marketing of consumer goods suggests an affirmative answer. Businessmen are taught that when selling a targeted product, they can enhance its appeal by positioning a clearly inferior product at nearly the same price (Simonson and Tversky, 1992). Thus, awareness illusions appear to be present in real markets, and systematic enough to be exploited by sellers.

Economists investigating consumer behavior can learn a great deal from careful study of market research findings and marketing practice. Ultimately, behavioral economists need to move beyond stylized descriptions of choice behavior and become involved in market research experiments that explore

Table 6. Willingness-to-Pay to Save 50,000 Off-Shore Seabirds per Year

Distribution	Open-Ended	Starting Point Bid				
		$5	$25	$60	$150	$400
$0-4.99	19.8%	12.2%	8.5%	0.0%	8.3%	12.0%
$5-24.99	27.3%	67.4%	25.5%	41.7%	29.2%	22.0%
$25-59.99	31.4%	12.2%	53.2%	14.6%	27.1%	20.0%
$60-149.99	12.4%	8.2%	8.5%	41.7%	16.7%	18.0%
$150-399.99	5.0%	0.0%	2.1%	2.1%	18.8%	10.0%
$400+	4.1%	0.0%	2.1%	0.0%	0.0%	18.1%
Sample size	121	49	47	48	48	50
P(Open-Ended Response>Bid)		80.2%	52.9%	21.5%	9.1%	4.1%
(Std. Error)		*(5.7%)*	*(7.1%)*	*(5.9%)*	*(4.1%)*	*(2.8%)*
P(Anchored Response>Bid)		87.8%	66.0%	43.8%	18.8%	18.0%
(Std. Error)		*(4.7%)*	*(6.9%)*	*(7.2%)*	*(5.6%)*	*(5.4%)*
Median Response	$25.00	$10.00	$25.00	$25.00	$43.00	$50.00
(Std. Error)	*($6.03)*	*($2.33)*	*($1.16)*	*($14.04)*	*($10.87)*	*($23.41)*
Mean Response (a)	$64.25	$20.30	$45.43	$49.42	$60.23	$143.12
(Std. Error)	*($13.22)*	*($3.64)*	*($12.61)*	*($6.51)*	*($8.59)*	*($28.28)*

	Coefficient	Std. Error
Marginal effect of starting point bid	0.284	0.32
Nonparametric referendum mean (b)	$167.33	$76.90
Parametric referendum mean	$265.59	$138.96

a. One observation of $2,000,000 is excluded from the calculation of the open-ended mean. If the open-ended mean WTP of $64.25 is representative of all California adults, then the total state WTP for protecting 50,000 seabirds is $1.49 bil., or $29,800 per bird.

b. The upper bound to the distribution is assumed to equal the largest anchored response, $1000. The reported std. error is the RMSE at the maximum possible bias, given the upper bound to the distribution.

Adapted from "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," D. McFadden with D. Green, K. Jacowitz, and D. Kahneman, Resource and Energy Economics, Vol. 20, 85–116, 1998.

directly the nature of economic choice processes. There may be a further methodological lesson from market research. Discovery and exploitation of cognitive illusions in purchase behavior seems to coexist comfortably with the use of RUM-consistent discrete response models, adapted to use data on perceptions, as a major tool for predicting buyer behavior.

V. STATISTICAL METHODS

The microeconometric analysis of choice behavior requires statistical methods for parametric and non-parametric estimation, and diagnostic tools to detect errors in specification and test hypotheses. Applications of choice models also require systems for producing disaggregate and aggregate forecasts and policy scenarios that track statistical accuracy. These requirements are generic to applied statistics, but are made more difficult in this area because natural models derived from RUM foundations are usually nonlinear, and often not particularly tractable.

Applied RUM analysis, based on the MNL model and its relatives, has generally relied on maximum likelihood methods and their large sample properties, and routines available in standard statistical software packages now permit more or less mindless use of these models. There is increasing use of non-parametric estimators, bootstrap methods to refine asymptotic approximations, generalized method of moments procedures for robustness, and simulation methods to overcome problems that are intractable using conventional computation. There are a few statistical developments that are specific to or particularly applicable to discrete choice analysis. I will summarize a few of these developments, concentrating on those in which I have had some hand.

A. Choice-Based Sampling

A *choice-based sample* is one obtained by stratification on the basis of response behavior whose explanation is the target of study. Observations on response and explanatory variables (covariates) are collected within each stratum. These are then used for statistical inference on the conditional distribution of the response, given the covariates. For example, a study of occupational choice may draw a sample stratified by occupation, so the first stratum is a sample of engineers, the second stratum is a sample of educators, and so forth. Data are collected on covariates such as gender and utilization of training subsidies. The observations might then be used to infer the impact of training subsidies on occupational choice. Choice-based samples may be unintentional, the result of self-selection or stratification in general purpose surveys, or may be deliberate, designed to reduce sampling costs or improve the informativeness or accuracy of responses.

Statistical methods developed for random samples will often be inconsistent or inefficient when applied to choice-based samples. The essential problem is that the analysis is attempting to infer properties of the conditional distribution of choices given covariates, using observations that are drawn from the conditional distribution of covariates given choices. The solution to the inference problem is to incorporate the mapping between the conditional distributions in the analysis, either by re-weighting observations so that they behave *as if* they were drawn from a random sample, or by re-weighting the probability model for a random sample so that it is consistent with the empirical sampling process. The statistical issues in analyzing choice based samples were treated in a seminal paper by Manski and Lerman (1977), with further results by Manski and McFadden (1981) and Steve Cosslett (1981). The choice-based sampling problem is closely related to the problem of analysis of self-selected samples. The seminal treatment of selection problems was given by Heckman (1974,1979), with further contributions by Hausman and Wise (1977), Goldfeld and Quandt (1973), Madalla and Nelson (1975), and Lung-

Fei Lee and Porter (1984). Papers by Hsieh, Manski, and McFadden (1985), Lancaster and Imbens (1990), and Breslow (1996) sharpen the statistical analysis of endogenous samples. Extensions of the basic framework for inference in choice-based samples have been made for a variety of problems. Imbens (1992) has provided methods for combining choice-based survey data with aggregate statistics. McFadden (1998) has studied the problem of analysis of endogenously recruited panels.

Table 7. Population Cell Probabilities

	y_1	y_2	.....	y_J	Sum
z_1	$P(y_1\|z_1)p(z_1)$	$P(y_2\|z_1)p(z_1)$	.....	$P(y_J\|z_1)p(z_1)$	$p(z_1)$
z_2	$P(y_1\|z_2)p(z_2)$	$P(y_2\|z_2)p(z_2)$	.....	$P(y_J\|z_2)p(z_2)$	$p(z_2)$
:	:	:		:	:
z_K	$P(y_1\|z_K)p(z_K)$	$P(y_2\|z_K)p(z_K)$	.....	$P(y_J\|z_K)p(z_K)$	$p(z_K)$
Sum	$q(y_1)$	$q(y_2)$	.....	$q(y_J)$	1

Table 7 depicts, schematically, the population probability law for a choice y and a vector of covariates z.[4] The joint probability of a (y,z) cell can be written as the product of the conditional probability of y given z times the marginal probability of z, $p(y,z) = P(y|z)\,p(z)$. The row sums give the marginal probability $p(z)$ of z, and the column sums give the marginal probability $q(y) = \Sigma_z P(y|z)\,p(z)$ of y. Bayes law gives the conditional probability of **z** given y, $Q(z|y) = P(y|z)p(z)/q(y)$. The target of statistical analysis is the conditional probability $P(y|z)$, sometimes termed the *response probability*. In applications, $P(y|z)$ is usually assumed to be *invariant under treatments* that alter the marginal probability of z; then knowledge of $P(y|z)$ permits the analysis to forecast y in new populations or under policy treatments where the z distribution is changed.[5]

Random sampling draws from the table in proportion to the cell probabilities. Exogenous stratification draws rows, with probabilities that may differ from the population marginal probabilities $p(z)$, and then within a row draws columns in proportion to their population conditional probabilities $P(y|z)$. A simple choice-based sampling design draws columns, with probabilities that may differ from the population marginal probabilities $q(y)$, then within a column draws rows in proportion to their conditional probabilities $Q(z|y) = P(y|z)p(z)/q(y)$.

[4] This exposition treats y and z as discrete, but the discussion applies with minor modifications to the case where y and/or some components of z are continuous.

[5] A conditional probability with this invariance property is sometimes said to define a causal model. It is true that a causal structure will imply this invariance property, but it is also possible for the invariance property to hold, making forecasting possible, without the presence of a deeper causal structure. Further, there are straightforward statistical tests for the invariance property, while detection of true causal structures is beyond the reach of statistics. For these reasons, it is best to avoid the language of causality and concentrate instead on invariance properties.

More complex endogenous sampling designs are also possible. A general framework that permits a unified analysis of many sampling schemes characterizes the sampling protocol for a stratum s in terms of a probability $R(z,y,s)$ that a member of the population in cell (z,y) will qualify for the stratum. The joint probability that a member of the population is in cell (z,y) and qualifies for stratum s is $R(z,y,s)\,P(y\,|z)\,p(z)$. The proportion of the population qualifying into the stratum, or *qualification factor,* is $r(s) = \Sigma_z\,\Sigma_y\,R(z,y,s)\,P(y\,|z)\,p(z)$, and the conditional probability of (z,y) given qualification is $R(z,y,s)\;P(y\,|z)\;p(z)/r(s)$. The term $R(z,y,s)$ is sometimes called a *propensity score.* When a fraction of the sample $f(s)$ is drawn from stratum s, $g(y,z) \equiv \Sigma_s\;R(z,y,s)\;P(y\,|z)\;p(z)f(s)/r(s)$ is the probability for an observation from the pooled sample, and $g(y\,|z) = P(y\,|z)(\Sigma_s R(z,y,s)\;f(s)/r(s))/[\Sigma_{y'} P(y'\,|z)\;(\Sigma_s R(z,y.,s)\;f(s)\;/r(s))]$ is the conditional probability of y given z in this pooled sample. Note that this conditional probability depends on the marginal probability of z only through the qualification factors.

When the sampling protocol is exogenous (i.e., $R(z,y,s)$ does not depend on y), the conditional probability $g(y\,|z)$ for the pooled sample equals the population conditional probability $P(y\,|z)$. Consequently, any statistical inference procedure designed to reveal features of the conditional probability $P(y\,|z)$ in random samples will apply to an exogenously stratified sample. In particular, if $P(y|z)$ is in a parametric family, then maximization of the random sample likelihood function in an exogenously stratified sample will have the same properties as in a random sample. However, in an endogenous sample in which the qualification probability $R(z,y,s)$ does depend on y, the conditional probability $g(y\,|z)$ for the pooled sample is *not* equal to $P(y\,|z)$. Consequently, statistical inference assuming that the data generation process is described by $P(y\,|z)$ is generally statistically inconsistent. Also, the distribution of covariates in an endogenous sample will differ from their population distribution, with $g(z) = p(z)\,\Sigma_s\;(f(s)/r(s))\Sigma_y\;R(z,y,s)P(yv|z)$, and a corresponding correction factor must be applied to the sample empirical distribution of z to estimate population quantities consistently.

Manski and McFadden (1981) propose that statistical inference when $P(y\,|z)$ is parametric be based on the conditional likelihood $g(v\,|z)$, and term this the *conditional maximum likelihood* (CML) method. When the qualification factors $r(s)$ and sample frequencies $f(s)$ are known or can be estimated consistently from external samples, and the forms of $P(y\,|z)$ and $R(z,y,s)$ allow identification of any unknown parameters in $R(z,y,s)$, this approach is consistent. In general, the probability $g(y\,|z)$ is not in the same parametric family as $P(v\,|z)$. To illustrate, suppose a population has a binomial probit choice probability, $P(2\,|z) = \Phi(\alpha + z\beta)$, and $P(1\,|z) = \Phi(-\alpha\text{-}z\beta)$. Suppose the sample consists of a randomly sampled stratum 1 with $R(z,y,1) \equiv 1$, plus a stratum 2 drawn from the population with response $y = 2$, with $R(z,y,2)$ equal to one if $y = 2$, and zero otherwise. This is called an *enriched sample.* The qualification factors are $r(1) = 1$ and $r(2) = q(2)$. If $q(2)$ is known, a consistent estimate of the slope parameter β in the model can be obtained by the CML method with

$$g(1\,|z) = \Phi(-\alpha\text{-}z\beta)f(1)/[\Phi(-\alpha\text{-}z\beta)\mathrm{f}(1) + \Phi(\alpha + z\beta)(f(1) + f(2)/q(2))].$$

By contrast, likelihood maximization using $P(y|z)$ is not consistent for β.

An important simplification of the CML method occurs for the MNL model. Suppose that the vector of covariates is partitioned into components $z = (v,x)$ with v discrete, and $P(y|v,x) = exp(\alpha_y + \gamma_{yv} + x\beta_y)/\Sigma_{y'} exp(\alpha_{y'} + \gamma_{y'v} + x\beta_{y'})$. In this model, the β_y are slope coefficients for the covariates x, the α_y are response-specific effects, and the γ_{yv} are interactions of response-specific and v-specific effects. Suppose that the qualification probability $R(v,x,y,s)$ does not depend on x. The conditional probability $g(y|z)$ is again of multinomial logit form, with the same β_y parameters but with the remaining parameters shifted; e.g., $g(y|v,x) = exp(\alpha^*_y + \gamma^*_{yv} + x\beta_y)/\Sigma_{y'} exp(\alpha^*_{y'} + \gamma^*_{y'v} + x\beta_{y'})$, with the transformed parameters satisfying $\alpha^*_y + \gamma^*_{yv} = \alpha_y + \gamma_{yv} + log(\Sigma_s R(v,y,s)\ f(s)/r(s))$. Consistent estimation of this model requires the inclusion of all the alternative specific effects and interactions that are modified by sampling factors. However, if these variables are included, then the slope parameters β_y are estimated consistently without further adjustments for endogenous sampling.[6]

B. Computation and Simulation

From an era where estimation of a single multinomial logit model was a major computational task, we have progressed to the point where simple multinomial logits are virtually instantaneous, even for large 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. However, the evaluation of choice probabilities that cannot be expressed in closed form, but require numerical integration of moderately high dimension, remains a computationally hard problem. For example, the multinomial probit model with an unrestricted covariance structure continues to resist conventional computation except for special cases.

Use of simulation methods has provided the most traction in obtaining practical representations and estimates for these computationally hard models. A simulated sample drawn from a trial data generation process (DGP) is an analog of a real sample drawn from the true DGP. If the simulation procedure is designed so that the simulated sample does not "chatter" as one varies the trial parameters, then one can estimate the true DGP by making the simulated and real samples congruent. McFadden (1989) develops and formalizes this approach to inference, and generalizes simulators for the multinomial probit model first introduced by Manski and Lerman (1981). Research in the past decade has expanded the library of simulation methods, 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 Method of

[6] Some statistical procedures use propensity score weights to remove correlation of treatment variables and covariates induced by exogenous self-selection.

Simulated Moments, Method of Simulated Scores, and the simulated EM algorithm (McFadden and Ruud, 1994; Hajivassiliou and McFadden, 1998; Hajivassiliou, McFadden, and Ruud 1996; Hajivassiliou and Ruud, 1994; Bhat, 2000; Train, 1999). These methods have made it feasible to work with quite flexible models, such as multinomial probit and mixed multinomial logit models. Statistical simulation is also a powerful tool for model comparison and policy analysis (Cowing and McFadden, 1984; Gourieroux and Monfort, 1996; Hendry, 1984). Considerable room for improvement in simulation methods remains. In particular, some of the statistical methods for dealing with measurement error and outliers in real data are also potentially useful for processing simulated data.

A model where simulation methods are usually needed, and relatively easy to apply, is the MMNL model (4). Under the name *kernel logit*, it has been employed by (McFadden, 1989; Bolduc, 1992; Brownstone and Train, 1999; Srinivasian and Mahmassani, 2000) as a computational approximation to multinomial probit or as a general flexible RUM approximation. Because the MNL model itself is smooth in its parameters α, the following procedure gives positive, unbiased, smooth simulators of the MMNL probabilities, and smooth simulators of their derivatives: Suppose α is given by a smooth parametric inverse mapping $\alpha(\varepsilon,\theta)$, 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 approximation $P_C(i) = E_R\, exp(Z(a\text{-}c_i, w, t_i, x_i, s)\, \alpha(\varepsilon_r,\theta))/\Sigma_j exp(Z(a\text{-}c_j, x_j, s)\, \alpha(\varepsilon_r,\theta))$, where E_R denotes an empirical expectation with respect to the simulation sample, as if it were exact. 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 this formula will contain a negligible simulation error in sufficiently large samples. To avoid misleading estimates of precision when sample sizes and R are moderate, one should use the sandwich formula for the covariance matrix in possibly misspecified models (McFadden and Train, 2000). In applications where the inverse transformation $\alpha(\varepsilon,\theta)$ is not tractable, one can instead use importance sampling methods or a Metropolis-Hastings sampler.

C. Specification Testing: IIA Tests

The MNL model, is a powerful tool for analysis of economic choice behavior when its 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 (McFadden, 1981, Atherton, Ben-Akiva, McFadden and 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 can 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 (McFadden, Tye, and Train, 1978; Hausman and McFadden 1984) required estimating 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 and A corresponds to a nest of similar alternatives, then there will 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. Hausman and McFadden showed that the quadratic form $(\beta_C-\beta_A)'(\Omega_A-\Omega_C)^{-1}(\beta_C-\beta_A)$ has an asymptotic 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 $\Omega_A-\Omega_C$. The simple form of the covariance matrix for the parameter difference arises because β_C is the efficient estimator for the problem.

Another test which is particularly easy to compute was proposed by McFadden (1987). Estimate the basic MNL model, using all the observations. Suppose A is a specified subset of alternatives. Create a new variable z_i that is zero for $i \in A$, and for $i \in A$ equals $log(P_A(i)) - \Sigma_{j\in A}P_A(j)\ log(P_A(j))$, where $P_A(j)$ is calculated from the basic model. A numerically equivalent form is obtained by replacing $log(P_A(j))$ by $V_j = x_j\,\beta$. Estimate an expanded MNL model that contains the basic model variables plus one or more of the new variables z_i constructed for different A. The A's can be disjoint, overlapping, and/or nested. Then carry out a likelihood ratio test for significance of the z's, with degrees of freedom equal to the number of added variables after eliminating any that are linearly dependent. If there is a single z, then the test can use the T-statistic for the significance of this variable. This test is asymptotically equivalent to a *score* or *Lagrange Multiplier* test of the basic MNL model against a *nested* MNL model in which consumers discriminate more sharply between alternatives within A than they do between alternatives that are not both in A. One minus the coefficient of a z variable can be interpreted as a preliminary estimate of the inclusive value coefficient for the nest A.

The test above for a single set A is asymptotically equivalent to a one-degree-of-freedom Hausman-McFadden test focused in the direction determined by the parameters β; conversely, the test above with the variable V_j replaced by the vector x_j for $j \in A$ is asymptotically equivalent to the original Hausman-McFadden test for A. One may get a rejection of the null hypothesis that IIA holds *either* if IIA is in fact false, *or* if there is some other problem with the model specification, such as omitted variables or a failure of the logit form due to asymmetry or fat tails in the disturbances. Rejection of the test will often occur when IIA is false even if the set A does not correspond to the true nesting pattern. However, the test will typically have greatest power when A is a nest for which an IIA failure occurs.

D. Specification Testing: Mixing in MNL Models

In light of the theoretical result that any well-behaved RUM model can be approximated by a MMNL 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 base MNL models, so that simulation estimators are not needed, and that it can test against a battery of alternatives at the same time. To perform the test, first construct artificial variables $z_{ti} = (x_{ti} - x_{tC})^2/2$ with $x_{tC} = \Sigma_{j\in C}\, x_{tj} \cdot P_C(j)$ for selected components t of x_i, where $P_C(j)$ are the estimated base MNL probabilities. Then re-estimate the model with these added variables and use a Wald or Likelihood Ratio test for the significance of the artificial variables. This test is asymptotically equivalent to a Lagrange multiplier test of the hypothesis of no mixing against the alternative of a MMNL model with mixing in the selected components t of the logit model. The degrees of freedom equals the number of artificial variables z_{ti} that are linearly independent of x. McFadden and Train (2000) 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.

E. Market Research Data and Models

An important inter-disciplinary interaction has developed between economic choice analysis and market research. The experimental methods used in market research permit elucidation and measurement of the workings of the decision-making process described in Figure 2. In particular, it is possible to elicit *stated perceptions*, *stated preferences*, and attitude scales; we call these *stated preference* (SP) data in contrast to the *revealed preference* (RP) data obtained from observed choices. Most of these variables and the methods used to measure them come from applied psychology. In particular, *conjoint analysis*, a method for eliciting stated preferences within a classical experimental design, provides data that with proper consumer training and allowance for cognitive illusions can add considerably to the ability to predict consumer decisions.

The use of experiments rather than field surveys to collect data on consumer decisions 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, including dimensions along which RP data provides no variation. Large quantities of relevant data can be collected at moderate cost. There will always be questions about how closely cognitive tasks in a hypothetical setting can match those in a real decision-making environment. Good experimental technique can remove the most obvious sources of incongruity, but calibration and validation using RP data is usually needed.

Both marketing and economic policy applications need an analytic framework for combining RP and SP data, and linking experience and information to SP responses. In 1984, Ben-Akiva and I specialized the multiple-indicator, multiple-cause (MIMC) model for this purpose, following the path diagram in Figure 2, and adding a hidden (latent) layer to handle mappings into discrete responses (Jöreskog and Sörbom, 1979; McFadden, 1986; Train, McFadden and Goett, 1987; Morikawa, 1989; Ben-Akiva and Morikawa, 1990). Applications have shown this to be useful framework for integrating marketing data into forecasting problems (Morikawa, Ben-Akiva, and Yamada, 1991; Louviere *et al.*, 1999; Hensher, Louviere, and Swait, 1989; Brownstone and Train, 1999).

VI. CONCLUSIONS

Looking back at the development of discrete choice analysis based on the RUM hypothesis, I believe that it has been successful because it emphasized empirical tractability and could address a broad array of policy questions within a framework that allowed results to be linked back to the economic theory of consumer behavior. Some possibilities for development of the approach have not yet been realized. The RUM foundation for applied choice 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 economic decisions along different dimensions and in different areas. The potentially important role of perceptions, ranging from classical psychophysical perception of attributes, through psychological shaping of perceptions to reduce dissonance, to mental accounting for times and costs, remains largely unexplored in empirical research on economic choice. Finally, the feedback from the empirical study of choice behavior to the economic theory of the consumer has begun, through behavioral and experimental economics, but is still in its adolescence.

What lies ahead? I believe that the basic RUM theory of decision-making, with a much larger role for experience and information in the formation of perceptions and expression of preferences, and allowance for the use of rules as agents for preferences, can describe most economic choice behavior in markets, surveys, and the laboratory. If so, then this framework can continue for the foreseeable future to form a basis for microeconometric analysis of consumer behavior and the consequences of economic policy.

REFERENCES

Ai, Chunrong and Chen, Xiaohong. "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknowns," London School of Economics Working Paper, 1999.

Anastassiou, George and Yu, Xiang."Convex and Coconvex Probabilistic Wavelet Approximation," STOCHASTIC ANALYSIS AND APPLICATIONS, 10, 1992, 507–521.
Atherton, Terry; Ben-Akiva, Moshe; McFadden, Daniel; and Train, Kenneth."Micro-simulation of Local Residential Telephone Demand Under Alternative Service Options and Rate Structures," in de Fontenay, Alain; Shugard, Mary and Sibley, David, eds., TELECOMMUNICATIONS DEMAND MODELING, Elsevier: Amsterdam., 1990, 137–163.
Ben-Akiva, Moshe. THE STRUCTURE OF TRAVEL DEMAND MODELS, Transportation Systems Division, Department of Civil Engineering, M.I.T., Ph.D. dissertation, 1972.
Ben-Akiva. Moshe and Morikawa, Takayuki. "Estimation of Switching Models from Revealed Preferences and Stated Intentions," TRANSPORTATION RESEARCH, 24A, 1990, 485–495.
Ben-Akiva, Moshe ; McFadden, Daniel; Gopinah, Dinesh; Garling, Tommy; Bolduc, Denis; Borsch-Supan, Axel; Delquie, Philippe, Larichev Oleg., Morikawa, Taka; Polydoropoulou, Amalia; Rao Vithala. "Extended Framework for Modeling Choice Behavior," MARKETING LETTERS, 10, 1999, 187–203.
Bhat, C. R. "Accommodating Flexible Substitution Patterns in Multidimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice," TRANSPORTATION RESEARCH 32B, 1998, 425–440.
Bhat, C. R. "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," Working Paper, Department of Civil Engineering, University of Texas, Austin, 2000.
Bolduc, Dennis. "Generalized Autoregressive Errors in the Multinomial Probit Model," TRANSPORTATION RESEARCH-B, 26B, 1992, 155–170.
Breslow, Norman. "Statistics in Epidemiology: The Case-control Study,"JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 91, 1996, 14–28.
Brown, Donald and Matzkin, Rosa. "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand." Yale Cowles Foundation Discussion Paper 1175, 1998.
Brownstone, David and Train, Kenneth "Forecasting New Product Penetration with Flexible Substitution Patterns," JOURNAL OF ECONOMETRICS, 89, 1999, 109–129.
Cardell, Scott and Dunbar, Fred. "Measuring the Societal Impacts of Automobile Downsizing," TRANSPORTATION RESEARCH, 14A, No. 5–6, 1980, 423–434.
Chen, Xiaohong and White, Hal. "Improved Rates and Asymptotic Normality in Nonparametric Neural Network Estimators." Working Paper, University of California, San Diego, 1998.
Cheng, Bing and Titterington, D. Michael. "Neural Networks: A Review from a Statistical Perspective," STATISTICAL SCIENCE, 9, 1994, 2–54.
Chipman, John. "The Foundations of Utility," ECONOMETRICA, 28, 1960, 193–224.
Chipman, John and Moore, James."Acceptable Indicators of Welfare Change," in John Chipman, Daniel McFadden, and James Moore, eds, PREFERENCES, UNCERTAINTY, AND OPTIMALITY, Westview Press, 1990.
Cosslett, Steve. "Maximum Likelihood Estimation for Choice-Based Samples," ECONOMETRICA, 49, 1981, 1289–1316.
Cottingham, Phoebe. MEASUREMENT OF NON-USER BENEFITS. Ph.D. dissertation, Dept. of Economics, Univ. of California, Berkeley, 1966.
Cowing, Thomas and McFadden, Daniel. MICROECONOMIC MODELING AND POLICY ANALYSIS, Academic New York, 1984.
Dagsvik, John. "Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Alternatives," ECONOMETRICA, 62, 1994, 1179–1205.
Daly, Andrew; Zachary, Stan. "Improved Multiple Choice Models," in Hensher, David and Dalvi, Quasim, eds., IDENTIFYING AND MEASURING THE DETERMINANTS OF MODE CHOICE, Teakfield: London, 1979.
Debreu, Gerard. "Review of R. D. Luce *Individual Choice Behavior*," AMERICAN ECONOMIC REVIEW, 50, 1960, 186–188.

Dechevsky, Lubomir and Penev, Spridon. "On Shape-Preserving Probabilistic Wavelet Approximations," STOCHASTIC ANALYSIS AND APPLICATIONS, 15, 1997, 187–215.
Domencich, Thomas and McFadden, Daniel. URBAN TRAVEL DEMAND, North Holland: Amsterdam., 1975.
Dubin, Jeffrey; McFadden, Daniel. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," ECONOMETRICA; 52, 1984, 345–62.
Fuss, Melvyn; McFadden, Daniel; and Mundlak, Yair "A Survey of Functional Forms in the Economic Analysis of Production," in Fuss, Melvyn and McFadden, Daniel, eds., PRODUCTION ECONOMICS, Vol. 1, 1978, 219–286.
Garling, Tommy "The Importance of Routines for the Performance of Everyday Activities," SCANDINAVIAN JOURNAL OF PSYCHOLOGY 33, 1992, 170–177.
Gourieroux, Christian and Monfort, Alain. SIMULATION-BASED ECONOMETRIC METHODS. Oxford: Oxford University Press, 1996.
Green, Donald; Jacowitz, Karen; Kahneman, Danny; and McFadden, Daniel. "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods," RESOURCE AND ENERGY ECONOMICS, 20, 1998, 85–116.
Griliches, Zvi. "Specification Bias in Estimates of Production Functions," JOURNAL OF FARM ECONOMICS, 39, 1957, 8–20.
Griliches, Zvi. "Hedonic Price Indices for Automobiles," in THE PRICE STATISTICS OF THE FEDERAL GOVERNMENT, NBER: Washington, 1961, 173–196.
Griliches, Zvi;.Ringstad, Vidar. ECONOMIES OF SCALE AND THE FORM OF THE PRODUCTION FUNCTION, North Holland: Amsterdam, 1970.
Goldfeld, Steve and Quandt; Richard "The Estimation of Structural Shifts by Switching Regressions," ANNALS OF ECONOMIC AND SOCIAL MEASUREMENT, 2, 1973, 475–486.
Hajivassiliou, Vassilis and Ruud, Paul. "Classical Estimation Methods for LDV Models using Simulation," in Robert Engle and Daniel McFadden, eds HANDBOOK OF ECONOMETRICS IV, North Holland: Amsterdam, 1994, 2384–2441.
Hajivassiliou, Vassilis; McFadden, Daniel; and Ruud, Paul. "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," JOURNAL OF ECONOMETRICS, 72, 1996, 85–134.
Hajivassiliou, Vassilis and McFadden, Daniel. "The Method of Simulated Scores with Application to Models of External Debt Crises," ECONOMETRICA, 66, 1998, 863–896.
Hausman, Jerry and Wise, David. "The Evaluation of Results from Truncated Samples: The New Jersey Negative Income Tax Experiment," ANNALS OF ECONOMIC AND SOCIAL MEASUREMENT, 5, 1976, 421–445.
Hausman, Jerry and McFadden, Daniel "Specification Tests for the Multinomial Logit Model," ECONOMETRICA, 52, 1984, 1219–1240.
Heckman, James. "Shadow Prices, Market Wages, and Labor Supply," ECONOMETRICA, 42, 1974, 679–694.
Heckman, James. "Sample Selection Bias as a Specification Error," ECONOMETRICA, 47, 1979, 153–171.
Heckman, James. "Statistical Models for the Analysis of Discrete Panel Data," in Manski, Charles and McFadden, Daniel, eds., STRUCTURAL ANALYSIS OF DISCRETE DATA, MIT Press, Cambridge, 1981a, 114–178.
Heckman, James. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Stochastic Process and Some Monte Carlo Evidence on Their Practical Importance," in Manski, Charles and McFadden, Daniel, eds., STRUCTURAL ANALYSIS OF DISCRETE DATA, MIT Press, Cambridge, 1981b, 179–185.
Hendry, David. "Monte Carlo Experiments in Econometrics," in Griliches, Zvi and Intrilligator, Michael, eds., HANDBOOK OF ECONOMETRICS,. II, 1396–1456, Elsevier: Amsterdam, 1984, 962–976.
Hensher, David.; Louviere, Jordan; and Swait, Jeffery. "Combining Sources of Preference Data," JOURNAL OF ECONOMETRICS, 87, 1999, 197–221.

Hsieh, David; Manski, Charles; and McFadden, Daniel. " Estimation of Response Probabilities from Augmented Retrospective Observations," JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION; 80, 1985, 651–62.

Hurd, Michael;. Merrill, Angela; and McFadden, Daniel. "Consumption and Savings Balances of the Elderly: Experimental Evidence on Survey Response," in David Wise, ed., FRONTIERS IN THE ECONOMICS OF AGING, 259–305, Univ. of Chicago Press: Chicago, 1998, 353–387.

Imbens, Guido. "An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling," ECONOMETRICA; 60, 1992, 1187–214.

Imbens, Guido and Lancaster, Tony. "Efficient Estimation and Stratified Sampling," JOURNAL OF ECONOMETRICS, 74, 1996, 289–318.

Jöreskog, Karl and Sörbom, Dag. ADVANCES IN FACTOR ANALYSIS AND STRUCTURAL EQUATION MODELS, Cambridge, Mass. Abt Books, 1979.

Kahneman, Danny and Tversky, Amos. "Intuitive Prediction: Biases and Corrective Procedures," STUDIES IN MANAGEMENT SCIENCE, 12, 1979, 313–327.

Kahneman, Danny and Tversky, Amos. "Choices, Values, and Frames," AMERICAN PSYCHOLOGIST 39, 1084, 341–350.

Karlstrom, Anders. "Non-linear Value Functions in Random Utility Econometrics," IATBR Conference, Australia, 2000.

Laffont, Jean Jacques and Quang Vuong, "Structural Analysis of Auction Data," AMERICAN ECONOMIC REVIEW, 86, 1996, 414–420.

Lancaster, Kevin. "A New Approach to Consumer Theory," JOURNAL OF POLITICAL ECONOMY, 74, 1966, 132–157.

Lancaster, Tony and Imbens, Guido. "Choice-Based Sampling of Dynamic Populations," in Hartog, Joop;Ridder, Geert; and Theeuwes, Jules, eds., PANEL DATA AND LABOR MARKET STUDIES, Amsterdam; North-Holland, 1990, 21–43.

Lee, Lung-Fei; Porter, Robert "Switching Regression Models with Imperfect Sample Separation Information, with an Application on Cartel Stability," ECONOMETRICA; 52, 1984, 391–418.

Louviere, Jordan; Meyer, Robert; Bunch, David; Carson, Richard; Dellaert, Benedict; Hanmeann, Michael; Hensher, David; Irwin, Julie. "Combining Sources of Preference Data for Modeling Complex Decision Processes," MARKETING LETTERS, 10, 1999, 205–217.

Lowenstein, George. "Out of control: Visceral Influences on Behavior," ORGANIZATIONAL BEHAVIOR AND DECISION PROCESSES, 65, 1996, 272–92.

Luce, R. Duncan. INDIVIDUAL CHOICE BEHAVIOR, Wiley: New York, 1959

Luce, R. Duncan; Suppes, Pat. "Preference, Utility, and Subjective Probability," in R. Luce, R. Bush & E. Galanter, eds., HANDBOOK OF MATHEMATICAL PSYCHOLOGY, Wiley: New York, 1965.

Maddala, G. S. and Nelson, Forrest. "Switching Regression Models with Exogenous and Endogenous Switching," PROCEEDINGS OF THE AMERICAN STATISTICAL ASSOCIATION (Business Section), 1975, 423–426.

Manski, Charles and Lerman, Steven. "The Estimation of Choice Probabilities from Choice Based Samples," ECONOMETRICA; 45, 1977, 1977–88.

Manski, Charles and Lerman; Steven. "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in Manski, Charles and McFadden, Daniel, eds., STRUCTURAL ANALYSIS OF DISCRETE DATA WITH ECONOMETRIC APPLICATIONS. 305–319, MIT Press: Cambridge, 1981, 305–319.

Manski, Charles;. McFadden, Daniel. "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in Manski, Charles and McFadden, Daniel, eds., STRUCTURAL ANALYSIS OF DISCRETE DATA WITH ECONOMETRIC APPLICATIONS. 305–319, MIT Press: Cambridge, 1981, 1981, 2–50.

Marschak, Jacob. "Binary Choice Constraints on Random Utility Indicators," Arrow, Kenneth, ed., STANFORD SYMPOSIUM ON MATHEMATICAL METHODS IN THE SOCIAL SCIENCES, Stanford Univ. Press: Stanford, 1960.

McFadden, Daniel. "The Revealed Preferences of a Public Bureaucracy," Dept. of Economics, Univ. of California, Berkeley, 1968.
McFadden, Daniel and Richter, Marcel K. "On the Extension of a Probability to the Boolean Algebra Generated by a Family of Events. " Working Paper, University of California, Berkeley, 1970.
McFadden, Daniel. "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka, Paul, ed., FRONTIERS IN ECONOMETRICS, Academic Press: New York, 1974a, 105–142.
McFadden, Daniel. "The Measurement of Urban Travel Demand," JOURNAL OF PUBLIC ECONOMICS, 3, 1974b, 303–328.
McFadden, Daniel. "The Revealed Preferences of a Government Bureaucracy: Empirical Evidence," THE BELL JOURNAL OF ECONOMICS AND MANAGEMENT SCIENCE, 7, 1976, 55–72.
McFadden, Daniel., Talvitie, Antti, and Associates. "Demand Model Estimation and Validation," URBAN TRAVEL DEMAND FORECASTING PROJECT, FINAL REPORT, VOLUME V, Institute of Transportation Studies, University of California, Berkeley, 1977.
McFadden, Daniel. "Estimation Techniques for the Elasticity of Substitution and Other Production Parameters," in Fuss, Melvyn and MxFadden, Daniel, eds., PRODUCTION ECONOMICS: A DUAL APPROACH TO THEORY AND APPLICATIONS, Vol 2, North-Holland: Amsterdam, 1978a, 73–124.
McFadden, Daniel."Modeling the Choice of Residential Location," in Karlqvist, Anders; Lundqvist, Lars; Snickars, Folke; and Weibull, Jorgen, eds., SPATIAL INTERACTION THEORY AND PLANNING MODELS, North Holland: Amsterdam, 1978b, 75–96.
McFadden, Daniel and Train, Kenneth. "The Goods/Leisure Tradeoff and Disaggregate Work Trip Mode Choice Models," TRANSPORTATION RESEARCH, 12, 1978, 349–353.
McFadden, Daniel; Tye, William; Train, Kenneth. "An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of the Multinomial Logit Model," TRANSPORTATION RESEARCH RECORD: FORECASTING PASSENGER AND FREIGHT TRAVEL, 637, 1978, 39–46.
McFadden, Daniel. "Econometric Models of Probabilistic Choice," in Manski , Charles and McFadden, Daniel, eds., STRUCTURAL ANALYSIS OF DISCRETE DATA WITH ECONOMETRIC APPLICATIONS, MIT Press: Cambridge, 1981, 198–272.
McFadden, Daniel. "Econometric Analysis of Qualitative Response Models," in Griliches, Zvi and Intrilligator, Mike, eds., HANDBOOK OF ECONOMETRICS,. II, Elsevier: Amsterdam, 1984, 1396–1456.
McFadden, Daniel."The Choice Theory Approach to Market Research," MARKETING SCIENCE, 5, 1986, 275–297.
McFadden, Daniel. "Regression-Based Specification Tests for the Multinomial Logit Model," JOURNAL OF ECONOMETRICS, 34, 1987, 63–82.
McFadden, Daniel. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," ECONOMETRICA, 57, 1989, 995–1026.
McFadden, Daniel and. Richter, Marcel K."Stochastic Rationality and Revealed Stochastic Preference." in Chipman, John; McFadden, Daniel; and Richter, Marcel K., eds., PREFERENCES, UNCERTAINTY, AND OPTIMALITY: ESSAYS IN HONOR OF LEO HURWICZ, 151–186, Westview Press: Boulder, 1990.
McFadden, Daniel and Ruud, Paul. "Estimation by Simulation." THE REVIEW OF ECONOMICS AND STATISTICS, 76, 591–608, 1994.
Hajivassiliou, Vassilis; McFadden, Daniel; and Ruud, Paul. "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives." JOURNAL OF ECONOMETRICS, 72, 85–134, 1996.
McFadden, Daniel. "On Endogenously Recruited Panels," 1998, JOURNAL OF APPLIED ECONOMETRICS, forthcoming.
McFadden, Daniel. "Rationality for Economists?" JOURNAL OF RISK AND UNCERTAINTY, 19, 1999a, 73–105.

McFadden, Daniel. "Computing Willingness-to-Pay in Random Utility Models," in Moore, James; Reizman, Raymond; Melvin, James, eds., TRADE. THEORY AND ECONOMETRICS, Routledge: London, 1999b, 253–274.

McFadden, Daniel and Train, Kenneth. "Mixed MNL Models for Discrete Response," 2000, JOURNAL OF APPLIED ECONOMETRICS, forthcoming.

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

Morikawa, Takayuki; Ben-Akiva, Moshe; and Yamada, K. "Forecasting Intercity Rail Ridership Using Revealed Preference and Stated Preference Data," TRANSPORTATION RESEARCH RECORD, 1328, 1991, 30–35.

Mundlak, Yair. "Estimation of Production and Behavioral Functions from a Combination of Cross-Section and Time-Series Data," in Christ, Carl *et al.*, eds, MEASUREMENT IN ECONOMICS, Stanford University Press: Stanford, 1963.

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

Revelt, David and Train, Kenneth. "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," REVIEW OF ECONOMICS AND STATISTICS, 80, 1998, 1–11.

Rust, John. "Structural Estimation of Markov Decision Processes." in Engle, Robert and McFadden, Daniel, eds. HANDBOOK OF ECONOMETRICS, Vol. 4, Amsterdam: Elsevier-North Holland, 1994, 3081–3143.

Simonson, Itamar and Tversky, Amos. "Choice in Context: Tradeoff Contrast and Extremeness Aversion," JOURNAL OF MARKETING RESEARCH, 29, 1992, 281–295.

Small, Kenneth. "A Discrete Choice Model for Ordered Alternatives," ECONOMETRICA 55, 1987, 409–424.

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

Svenson, Ola. "Process Descriptions of Decision Making," ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE 23, 1979, 86–112.

Taussig, Frank. PRINCIPLES OF ECONOMICS, Macmillan: New York, 1912.

Thurstone, L. L "A Law of Comparative Judgment," PSYCHOLOGICAL REVIEW, 34, 1927, 273–286.

Train, Kenneth. "Recreation Demand Models with Taste Differences over People," LAND ECONOMICS, 74, 1998, 230–239.

Train, Kenneth. "Halton Sequences for Mixed Logit," Dept. of Economics, Univ. of California, Berkeley, 1999.

Train, Kenneth; McFadden; Daniel; and Goett, Andrew. "Consumer Attitudes and Voluntary Rate Schedules for Public Utilities," REVIEW-OF-ECONOMICS-AND-STATISTICS; 69, 1987, 383–91.

Tversky, Amos and Kahneman, Danny. "Judgment under uncertainty: heuristics and biases" SCIENCE, 185, 1974, 1124–1131.

Tversky, Amos and Kahneman, Danny. "The Framing of Decisions and the Psychology of Choice," SCIENCE, 211, 1981, 453–458.

White, Hal. "Some Asymptotic Results for Learning in Single Hidden Layer Feedforward Networks," Journal of the American Statistical Association, 84, 1989, 1008–1013.

White, Hal. ARTIFICIAL NEURAL NETWORKS: APPROXIMATION AND LEARNING THEORY, Blackwell: Oxford, 1992.

Williams, H. W. C. L. "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit," ENVIRONMENT & PLANNING, A9, 1977, 285–344.