RANDOM UTILITY MODELS WITH BOUNDED AMBIGUITY

Charles F. Manski
Department of Economics and Institute for Policy Research
Northwestern University

Draft: April 27, 2005

Prepared for presentation at a conference honoring Daniel McFadden, May 6-7, 2005. This research is supported in part by National Science Foundation grant SES-0314312. I am grateful to Maria Salgado for excellent research assistance.

1. Introduction

Econometric analysis of discrete choice has made considerable use of random utility models (RUMs) to interpret observed choice behavior (McFadden, 1974, 1981). Let J be a population of decision makers, each of whom chooses an action from a finite choice set C. The standard deterministic RUM assumes that person j associates utilities (u_{jc}, $c \in C$) with the feasible actions and chooses one that maximizes utility. The standard inference problem is to learn the distribution of preferences from data on the choices and covariates of a random sample of decision makers. Let X denote the feasible values of the observable covariates. Assume that the distribution of preferences is continuous and has the form F_θ, where θ belongs to a specified parameter space Θ. Then the equations

$$(1) \quad P(c|x) = F_\theta(u_c \geq u_d, d \in C|x), \qquad (c, x) \in C \times X$$

relate the choice probabilities $P(c|x)$ to the distribution of preferences F_θ. The identification region for θ is the set of parameter values that satisfy equations (1). The usual practice is to restrict the parameter space enough to point-identify θ and, hence, fully reveal the distribution of preferences.

Most empirical research today concerns choice problems in which decision makers act with partial knowledge of the utilities of the feasible actions. Economists use random expected utility models to analyze such choice problems. Let Γ be a set of states of nature and, for $\gamma \in \Gamma$, let $u_{jc\gamma}$ be the utility of action c to person j in state of nature γ. Suppose that, at the time of decision making, person j does not know what state of nature will be realized. Researchers routinely assume that person j places a subjective probability distribution on Γ, say Q_j, and chooses an action that maximizes expected utility. They assume that the joint distribution of preferences and expectations is continuous and has the form F_θ, where θ belongs to a specified parameter space Θ. Then the equations

$$(2)\quad P(c|x) \;=\; G_\theta(\textstyle\int u_{c\gamma}dQ \geq \int u_{d\gamma}dQ,\; d \in C|x), \qquad (c, x) \in C \times X$$

relate $P(c|x)$ to G_θ. The most common practice is to specify fully the expectations that persons are presumed to hold, in which case the task of choice analysis reduces to inference on preferences alone.

In Manski (1993, 2004a) and elsewhere, I have argued that the expectations assumptions made in empirical research rarely have much foundation. To enhance the credibility of econometric analysis, I have recommended direct measurement of the expectations that decision makers hold. But what can one do in the absence of expectations data? In this case, one can still study how inference depends on the expectations assumptions imposed. This paper shows how.

The idea, simply enough, is to specify a set of expectations that decision makers may plausibly hold, rather than make untenable assumptions that they hold particular expectations. Empirical research specifying a set of feasible expectations typically cannot point-identify the population distribution of preferences, but it can yield more credible partial-identification findings. It can also make plain the extent to which conventional point estimates rest on untenable expectations assumptions.

Section 2 poses the idea in abstraction and then specializes to the case of binary response with linear utilities, where the implications for inference are particularly straightforward. When the random expected utility model contains a single variable with uncertain value, the Manski and Tamer (2002) analysis of monotone index models with interval regressor data can be brought to bear. Their modified minimum distance estimator can be used to estimate consistently the identification region of the parameters. I illustrate using data on travel mode choice.

The analysis of Section 2 assumes that decision makers possess unique subjective probability distributions on the states of nature and make choices that maximize expected utility. Section 3 considers the possibility that persons place only partial probabilistic structure on the states of nature and make choices in some manner that uses the available structure. Although this situation is behaviorally distinct from the one considered earlier in the paper, the formal analysis of Section 2 still applies. I suggest that choice between an innovation and a status quo alternative is a common class of problem in which persons may not have complete probabilistic expectations.

I refer to the models of choice behavior developed in this paper as random utility models with bounded ambiguity (RUMBAs). In Section 2, the ambiguity is observational, in that the researcher does not observe the expectations that decision makers hold. In Section 3, the ambiguity is behavioral, in that persons do not possess complete probabilistic expectations. The adjective "bounded" refers to the fact that non-trivial inference on the population distribution of preferences is possible only if the researcher possesses some knowledge of their expectations.

2. Observational Ambiguity

2.1. Generalities

I assume throughout that the researcher and decision makers agree on the set Γ of states of nature. Let Ω_j denote the set of subjective distributions on Γ that the researcher deems plausible for person j to hold. I assume that the researcher is correct in thinking that the set Ω_j contains the person's expectations.

Let a_j denote the action chosen by person j. Whatever specific expectations person j may hold, he should not choose a_j if there exists another action that yields higher expected utility under all distributions in Ω_j. Hence,

$$(3a) \qquad a_j = c \;\Rightarrow\; \nexists\, d \in C \;\text{ s. t. }\; \int u_{jd\gamma} dQ > \int u_{jc\gamma} dQ,\ \forall\, Q \in \Omega_j.$$

If there exists an action c that outperforms all other actions under all distributions in Ω_j, person j should choose this action. Hence,

$$\text{(3b)} \quad \int u_{jc\gamma}dQ > \int u_{jd\gamma}dQ, \forall (d, Q) \in C \times \Omega_j \text{ s. t. } d \neq c \;\Rightarrow\; a_j = c.$$

Aggregating across the population, (3a) and (3b) imply these inequalities relating choice probabilities to the distribution of preferences and expectations:

$$\text{(4)} \quad G_\theta[\int u_{c\gamma}dQ > \int u_{d\gamma}dQ, \forall (d, Q) \in C \times \Omega \text{ s. t. } d \neq c|x] \;\leq\; P(c|x)$$

$$\leq G_\theta[\nexists\, d \in C \text{ s. t. } \int u_{d\gamma}dQ > \int u_{c\gamma}dQ, \forall Q \in \Omega|x], \quad (c, x) \in C \times X.$$

These inequalities provide the basis for inference on θ. The identification region is the set of parameter values that satisfy (4).

Inference with No Knowledge of Expectations

The worst case of partial knowledge of expectations occurs when, for each person j, Ω_j is the set of all probability distributions on Γ. Then the researcher knows nothing at all about the expectations that decision makers hold.

In this case, Ω_j includes the degenerate distributions that place all of their mass on one state of nature. The linearity of expected utilities in probabilities implies that if the logical relationship (3a)-(3b) hold for these degenerate distributions, then they hold for all elements of Ω_j. Hence, inequalities (4) reduce to

$$\text{(5)} \quad G_\theta[u_{c\gamma} > u_{d\gamma}, \forall (d, \gamma) \in C \times \Gamma \text{ s. t. } d \neq c|x] \;\leq\; P(c|x)$$

$$\leq G_\theta[\nexists\, d \in C \text{ s. t. } u_{d\gamma} > u_{c\gamma}, \forall \gamma \in \Gamma|x], \quad (c, x) \in C \times X.$$

2.2. Binary Choice with Linear Utilities

Although inequalities (4) describe the inferential problem in generality, they are too abstract to communicate much. Hence, I now consider the special case of binary choice with linear utilities.

Assume that each member of the population must choose between two actions, labeled 0 and 1. The utility of action c to person j is

$$(6) \qquad u_{jc\gamma} \equiv z_{jc}\beta + \alpha y_{jc\gamma} + \epsilon_{jc}.$$

Here z_{jc} is a K-vector, $y_{jc\gamma}$ and ϵ_{jc} are scalar, and (β, α) are corresponding parameters. The absence of a γ-subscript on (z_{jc}, ϵ_{jc}) indicates that the person knows these quantities at the time of decision making. The presence of a γ-subscript on $y_{jc\gamma}$ indicates that this quantity depends on the unknown state of nature. Assume that the person chooses an action that maximizes expected utility. Let $z_j \equiv z_{j1} - z_{j0}$, $y_j \equiv y_{j1} - y_{j0}$, $v_{j\gamma} \equiv \int y_{j\gamma} dQ_j$, and $\epsilon_j \equiv \epsilon_{j1} - \epsilon_{j0}$. Then the decision rule is

$$(7) \qquad a_j = 1[z_j\beta + \alpha v_j + \epsilon_j > 0].$$

Suppose that a researcher draws a random sample of N members of the population. For each sample member $j = 1, \ldots, N$, the researcher observes (a_j, z_j) but not (Q_j, ϵ_j). This would be a standard problem of binary choice analysis if the researcher were to observe Q_j, or at least v_j. However, the problem of interest is inference in the absence of data on the expectations that sample members hold. The prevailing practice is to assume that v_j takes some particular value and proceed with standard binary choice analysis. Here I assume instead that $Q_j \in \Omega_j$. I also assume that the sign of α is known; for convenience, the discussion below assumes that $\alpha \geq 0$.

Let $v_{j0} \equiv \inf(\int y_{j\gamma}dQ, Q \in \Omega_j)$ and $v_{j1} \equiv \sup(\int y_{j\gamma}dQ, Q \in \Omega_j)$. Given (6), (7), and $\alpha \geq 0$, the logical relationships in (3) reduce to

(8a) $\quad a_j = 1 \Rightarrow z_j\beta + \alpha v_{j1} + \epsilon_j > 0,$

(8b) $\quad z_j\beta + \alpha v_{j0} + \epsilon_j > 0 \Rightarrow a_j = 1.$

Aggregating across the population, (8a) and (8b) yield these inequalities relating choice probabilities to the distribution of preferences:

(9) $\quad G_\theta[z\beta + \alpha v_0 + \epsilon > 0|x] \leq P(1|x) \leq G_\theta[z\beta + \alpha v_1 + \epsilon > 0|x], \quad x \in X.$

The observed covariates x include (z_0, z_1, Ω). The parameter θ includes (β, α) plus the parameters needed to describe $P(\epsilon|x)$, the distribution of ϵ conditional on x. The identification region is the set of parameter values that satisfy (9).

The remainder of this section studies the structure of the identification region. I begin with an important negative finding. The identification region is generically unbounded if expectations have an indeterminate sign.

Unbounded Ambiguity when Expectations Have an Indeterminate Sign

Suppose that, for each person j, the researcher does not know enough about expectations to determine the sign of v_j; thus, $v_{j0} < 0 < v_{j1}$. Suppose that, for all feasible values of β and $P(\epsilon|x)$, the parameter space Θ places no upper bound on α. Let $\alpha \to \infty$. Then, for each person j, $z_j\beta + \alpha v_{j0} + \epsilon_j \to -\infty$ and $z_j\beta + \alpha v_{j1} + \epsilon_j \to \infty$. Hence, $\lim_{\alpha \to \infty} G_\theta[z\beta + \alpha v_0 + \epsilon > 0|x] = 0$ and $\lim_{\alpha \to \infty} G_\theta[z\beta + \alpha v_1 + \epsilon > 0|x] = 1$.

This result, combined with mild regularity conditions, implies that observation of choice behavior places no restrictions on β or $P(\epsilon|x)$ beyond those already present in the specification of the parameter space Θ. Consider any feasible value of $[\beta, P(\epsilon|x)]$. Holding these parameters fixed, let $\alpha \to \infty$. Suppose that

$$\text{(10a)} \quad \lim_{\alpha \to \infty} \sup_{x \in X} \{G_\theta[z\beta + \alpha v_0 + \epsilon > 0|x] - P(1|x)\} < 0,$$

$$\text{(10b)} \quad \lim_{\alpha \to \infty} \sup_{x \in X} \{G_\theta[z\beta + \alpha v_1 + \epsilon > 0|x] - P(1|x)\} > 0.$$

Then there exists an $\alpha^*[\beta, P(\epsilon|x)] > 0$ such that the inequalities (9) collectively hold for all $\alpha > \alpha^*[\beta, P(\epsilon|x)]$. Hence, the identification region contains the specified value of $\{\beta, P(\epsilon|x)]$ and all $\alpha > \alpha^*[\beta, P(\epsilon|x)]$.

The uniformity conditions (10a)-(10b) typically hold in practice. In particular, they hold if the covariate space X is finite and $0 < P(1|x) < 1$ for all $x \in X$. Thus, meaningful inference is possible only if the researcher knows the sign of $\int y_{j\gamma} dQ_j$ for at least some decision makers or can place a suitably tight upper bound on α.

Monotone-Index Models

Empirical research on binary choice often assumes that ϵ is statistically independent of x with a specified strictly increasing distribution function, such as the standard normal or logistical distribution. Let F be the assumed distribution function for $-\epsilon$. Then (9) becomes

$$\text{(11)} \quad F(z\beta + \alpha v_0) \leq P(1|x) \leq F(z\beta + \alpha v_1), \quad x \in X,$$

where the parameter θ is now the pair (β, α). The identification region for (β, α) is the set of parameter values that satisfy inequalities (11).

This is a monotone-index model with interval regressor data, of the form studied in Manski and

Tamer (2002, Section 4). Their Corollary to Proposition 4 shows that the identification region for (β, α) is convex. Given some regularity conditions, their *modified minimum-distance* (MMD) method provides a consistent estimate of this region.

Manski and Tamer (2002) also study the case in which the researcher knows only that some quantile of $P(\epsilon | x)$ is constant on X. The identification region remains convex, but larger than the one obtained with a monotone-index model. This region may be estimated using a *modified maximum score* method.

2.3. Empirical Illustration

The analysis of travel mode choice for the journey between home and work was one of the earliest applications of random utility models and has remained an important subject of empirical research (e.g., Warner, 1962; Domencich and McFadden, 1974). The canonical mode-choice model supposes that, each day, a worker chooses between two alternatives, travel by automobile and by public transit. The utility of each mode depends on the associated travel cost and travel time.

Empirical researchers have commonly used models of traffic flow on transportation networks to impute the travel times that particular workers would experience by each mode. Researchers have also assumed that these imputed travel times agree with the travel times that workers perceive when they make their mode choices. The accuracy with which imputed travel times measure travel time expectations is questionable. Transportation network models cannot precisely emulate the circumstances of individual travelers. Moreover, workers typically are uncertain how long the journey will take by each mode; travel times may vary from day to day due to unforeseen variation in traffic volume and the possibility of accidents.

To obtain a sense of how travel time imputations affect research findings, I use data on mode choice in the Washington, D.C. metropolitan area that have previously been analyzed in Horowitz (1993). In this illustration, mode 0 is public transit and mode 1 is automobile. The distribution function F is standard logistic. The vector z includes the four components (*constant*, *autos*, *dovtt*, *dcost*) and the variable y is *divtt*, where

constant = 1 always

autos = number of cars owned by the traveller's household

dovtt = automobile out-of-vehicle travel time minus transit out-of-vehicle travel time (minutes)

dcost = automobile travel cost minus transit fare (dollars)

divtt = automobile in-vehicle travel time minus transit in-vehicle travel time (minutes).

Thus, I suppose that the imputed values of travel cost and out-of-vehicle travel time match workers' perceptions of these variables. However, the imputed values of in-vehicle travel time need not agree with subjective expected values for in-vehicle travel time.

Table 1 gives descriptive statistics for the 842 observations in the Washington data. Taking the data at face value, maximum likelihood estimates for the coefficients are as follows:

$$\beta_{\text{constant}} = -1.222, \quad \beta_{\text{cars}} = 2.308, \quad \beta_{\text{dovtt}} = -0.062, \quad \beta_{\text{dcost}} = -0.017, \quad \alpha = -0.009.$$

These estimates displays the qualitative features typical of empirical analysis of mode choice, although the implied values of travel time are quite high relative to other studies. Ceteris paribus, the utility of each model decreases with it travel time and cost, with out-of-vehicle travel time being more onerous than in-vehicle travel time. The implied monetary value of out-of-vehicle and in-vehicle travel times are \$3.65 and \$0.53 per minute.

Table 2 reports modified minimum-distance estimates for three specifications of (v_0, v_1). In each case, I select a constant $\delta \geq 0$ and let $(v_0 = \text{divtt} - \delta, v_1 = \text{divtt} + \delta)$. Thus, each specification assumes that expected travel time lies in an interval of width 2δ, centered at the imputed travel time divtt. Estimation is performed under the assumption that $\alpha \leq 0$, which is natural in this setting.

The existing proof of consistency for the MMD estimator requires that the set-valued estimate include any coefficient value that makes the criterion function sufficiently close to its global minimum, where the researcher tightens the definition of "sufficiently close" as the sample size increases. Table 2 reports two sets of estimates, one using a relatively tight notion of closeness (up to 1.01 times the minimum) and the other using a much looser notion (up to 1.1 times the minimum). For each value of δ, the latter set-estimate necessarily encompasses the former one. The discussion below focuses on the former estimates.

The estimate setting $\delta = 0$ assumes that expected and imputed travel times coincide. In this case, the parameters are point-identified. Inclusion of coefficient values that are "sufficiently close" to the minimum still produces a set-estimate, but the resulting set is small. The overall impression is that the MMD set-estimate is reasonably similar to the maximum likelihood point estimate.

The estimates setting $\delta = 10$ and $\delta = 30$ permit expected travel times to differ from imputed ones by as much as 10 and 30 minutes respectively. The estimates for $\delta = 0$ and $\delta = 10$ are reasonably similar to one another. Thus, the possibility of smallish differences between expected and imputed travel times does not seriously degrade the quality of inference. However, the estimates for $\delta = 30$ differ substantially from those obtained with $\delta =10$. Indeed the interval-estimate for β_{dovtt} rises from [-0.030, -0.011] to [0.201, 0.253] and the one for α falls from [-0.040, -0.028] to [-0.240, -0.196].

The identification region for (β, α) grows as δ increases. Hence, one might anticipate that the interval-estimates for each parameter would be nested as one moves from the top to bottom panel of the table. The intervals do widen as δ increases but, as observed above for β_{dovtt} and α, they sometimes are not nested. Non-nesting with sample data is algebraically possible even if the logit specification of the choice model is correct. Non-nesting can also occur if the model is misspecified.

3. Behavioral Ambiguity

3.1. Generalities

In Section 2, I supposed that decision makers have complete subjective probability distributions on the states of nature, which they use to maximize expected utility. The inferential problem was observational, in that the researcher did not know what expectations people have. In this section, I suppose that person j wants to maximize expected utility, but does not have a unique subjective distribution on the states of nature. Instead, the person has a set Ω_j of such distributions, as in Gilboa and Schmeidler (1989), Walley (1991), and much other research on choice under ambiguity.

Decision theorists have suggested various criteria for choice under ambiguity, such as the maximin and minimax-regret rules and the class of Hurwizc rules. Although there is no consensus on how persons with incomplete probabilistic expectations should or do behave, decision theorists do largely agree that persons should not choose actions that are strictly dominated. In the present setting, this means that person j should not choose action a_j if there exists another action that yields higher expected utility under all distributions in Ω_j. Moreover, if there exists an action c that outperforms all other actions under all distributions in Ω_j, person j should choose this action.

These dominance conditions mean that the logical relationships (3a)-(3b) and the inequalities (4) hold when ambiguity is behavioral. Hence, a researcher who knows the set of distributions Ω_j held by each person j can analyze choice data in the same manner as discussed in Section 2. It may be that a researcher does not know Ω_j but can specify a larger set $\Omega_j' \supset \Omega_j$ that he believes to encompass Ω_j. Then the researcher may proceed with Ω_j', which expresses both behavioral and observational ambiguity.

3.2. Adoption of Innovations

Economists have long wanted to understand the manner in which decision makers learn about and choose innovations. A common scenario envisions an initial condition in which all decision makers choose an existing *status quo* action whose outcomes are known from historical experience. At some point, an innovation with unknown outcomes becomes available. From then on, successive cohorts of decision makers choose between the status quo and the innovation, with later cohorts observing the experiences of earlier ones and possibly learning from them.

Conventional econometric analysis of the adoption of innovations assumes that each decision maker has well-defined probabilistic expectations for the outcomes of the innovation and, moreover, specifies what those expectations are. However, in the absence of expectations data, researchers have little if any basis for making particular assumptions. Indeed, decision makers may not have complete probabilistic expectations.

When the innovation is introduced, the first cohort of decision makers who can choose this alternative have no historical experience to draw on. It seems plausible that members of this cohort makers may feel unable to place a complete subjective distribution on the outcomes that the innovation produces. Later cohorts can observe the experience of persons who chose the innovation and so may be willing to form more complete probabilistic expectations. Thus, ambiguity may lessen as time passes.

A specific model of this process is studied in Manski (2004b). There, I examine the dynamic process of information accumulation and decision making that occurs as the members of each cohort observe the experiences of earlier ones, and then make choices that yield experiences observable by future cohorts. I consider the extreme case in which members of the first cohort of decision makers have no probabilistic expectations at all. Subsequent cohorts can learn from the experiences of earlier ones, but learning is incomplete because all decision makers face a basic identification problem, the *selection problem*. The problem is that one can observe the outcomes of the innovation only for those earlier decision makers who chose the innovation, not for those who chose the status quo. As a consequence, I show that each decision maker j is able to place a particular set of distributions Ω_j on the outcomes of the innovation, the size and structure of Ω_j depending on the cohort to which j belongs.

References

Domencich, T. and D. McFadden (1975), *Urban Travel Demand: A Behavioral Analysis*, Amsterdam: North-Holland.

Gilboa, Y. and D. Schmeidler (1989), “Maxmin Expected Utility with a Non-Unique Prior,” *Journal of Mathematical Economics*, 18, 141-153.

Horowitz, J. (1993), “Semiparametric Estimation of a Work-Trip Mode Choice Model,” *Journal of Econometrics*, 58, 49-70.

Manski, C. (1993), “Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?” in C. Clotfelter and M. Rothschild (editors), *Studies of Supply and Demand in Higher Education*, Chicago: University of Chicago Press, 43-57.

Manski, C. (2004a), “Measuring Expectations,” *Econometrica*, 72, 1329-1376.

Manski, C. (2004b), “Social Learning from Private Experiences: The Dynamics of the Selection Problem,” *Review of Economic Studies*, 71, 443-458.

Manski, C. and E. Tamer (2002), “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70, 519-546.

McFadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka (editor), *Frontiers in Econometrics*, New York: Academic Press.

McFadden, D. (1981), “Econometric Models of Probabilistic Choice,” in C. Manski and D. McFadden (editors), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Mass.:M.I.T. Press.

Walley. P. (1991), *Statistical Reasoning with Imprecise Probabilities*, London: Chapman & Hall.

Warner, S. (1962), *Stochastic Choice of Mode in Urban Travel: a Study in Binary Choice*, Evanston: Northwestern University Press.

.

Table 1: Travel Mode Choice Data

	mode	cars	dovtt	dcost	divtt
minimum	0.00	0.00	-48.00	-89.00	-102.00
maximum	1.00	7.00	6.00	111.00	59.00
mean	0.84	1.50	-12.85	12.94	-17.05
std. dev.	0.37	0.87	10.06	37.97	17.96
median	1.00	1.00	-11.00	6.50	-13.00

Table 2: Modified Minimum-Distance Estimates of Travel Mode Choice Model

	$v_0 = v_1 = v$			
	1.01·(function minimum)		1.1·(function minimum)	
	lower bound	upper bound	lower bound	upper bound
constant	-2.010	-1.627	-2.558	-1.234
cars	3.077	3.519	2.600	4.203
dovtt	-0.056	-0.038	-0.078	-0.016
dcost	-0.020	-0.016	-0.026	-0.011
divtt	-0.018	-0.009	-0.030	-0.000
function minimum = 4.019				

	$v_0 = v - 10,\ v_1 = v + 10$			
	1.01·(function minimum)		1.1·(function minimum)	
	lower bound	upper bound	lower bound	upper bound
constant	-1.759	-1.396	-2.287	-1.101
cars	2.685	3.043	2.312	3.748
dovtt	-0.030	-0.011	-0.059	0.016
dcost	-0.021	-0.017	-0.028	-0.012
divtt	-0.040	-0.028	-0.057	-0.014
function minimum = 3.065				

	$v_0 = v - 30,\ v_1 = v + 30$			
	1.01·(function minimum)		1.1·(function minimum)	
	lower bound	upper bound	lower bound	upper bound
constant	-1.362	-1.012	-2.123	-0.381
cars	4.199	4.707	3.847	5.717
dovtt	0.201	0.253	0.146	0.369
dcost	-0.071	-0.059	-0.093	-0.051
divtt	-0.240	-0.196	-0.327	-0.166
function minimum = 0.114				