Copenhagen Business School
HANDELSHØJSKOLEN

Department of Economics

Copenhagen Business School

Working paper 4-2010

NON-LINEAR MIXED LOGIT

Steffen Andersen, Glenn W. Harrison,
Arne Risa Hole and E. Elisabet Rutström

Department of Economics -Porcelænshaven 16A, 1.fl. - DK-2000 Frederiksberg

Non-Linear Mixed Logit

by

Steffen Andersen, Glenn W. Harrison, Arne Risa Hole and E. Elisabet Rutström†

March 2010

Abstract. We develop an extension of the familiar linear mixed logit model to allow for the direct estimation of parametric non-linear functions defined over structural parameters. A classic application is the estimation of coefficients of utility functions to characterize risk attitudes. There are several unexpected benefits of this extension, apart from the ability to directly estimate structural parameters of theoretical interest.

† Department of Economics, Copenhagen Business School, Copenhagen, Denmark (Andersen); Department of Risk Management & Insurance and CEAR, Robinson College of Business, Georgia State University, USA (Harrison); Department of Economics, University of Sheffield, UK (Hole); and Department of Economics, College of Business Administration, University of Central Florida, USA (Rutström). E-mail contacts: sa.eco@cbs.dk, gharrison@gsu.edu, a.r.hole@sheffield.ac.uk and erutstrom@bus.ucf.edu. We thank the U.S. National Science Foundation for research support under grants NSF/HSD 0527675 and NSF/SES 0616746.

The linear mixed logit model provides a valuable inferential tool, allowing unobserved individual heterogeneity in the population to be characterized by random coefficients following some parametric distribution. There are several advantages, however, from considering the extension to the *parametric non-linear* mixed logit model.[1] First, and foremost, one can directly identify and estimate parameters from non-linear structural models, such as coefficients that reflect risk aversion in individual choice settings. Second, one can easily include non-linear transformations of familiar parametric families, such as the Normal, that allow theoretically constrained *and* flexible characterizations of population distributions. Third, one can directly estimate coefficients on demographic explanatory variables without having to "interact" them with task characteristics.

We state the formal non-linear mixed logit model in section 1, consider the immediate extension to allow flexible population distributions for the random parameters in section 2, and offer an illustrative applications in section 3. In section 4 we explain the differences between a parametric non-linear specification and semi-parametric and flexible-functional-form alternatives that have been proposed long ago in the econometric literature.

1. Linear and Non-Linear Mixed Logit

For pedagogic purposes, we consider the formulation of parametric, non-linear mixed logit models to estimate structural parameters of a decision-making model of risk attitudes. One natural reason for this focus is that mainstream theory *defines* risk attitudes in terms of the non-linearity of the utility function. Even non-mainstream utility theories typically allow for non-linearity in utility and/or probability weighting functions, so this application is arguably canonical for the estimation of structural parameters applied in non-linear functional forms. The methodology could just as easily be applied to recover structural parameters in production functions. The overall objective is to be able to recover estimates of the "deep" parameters of a structural theory, not some linear reduced

[1] Of course all logit models are non-linear in the sense that the probabilities are non-linear functions of the latent index, which is a linear function of the explanatory variables and the estimated coefficients. We use the term "non-linear mixed logit" to refer to mixed logit models with a latent index that is a non-linear function of these variables and coefficients.

form implication of that theory.

Assume a sample of N subjects making choices over J lotteries in T experimental tasks.[2] In all of the applications we consider, J=2 since the subjects are making choices over two lotteries, but there are many designs in which the subject is asked to make choices over J>2 lotteries (e.g., Binswanger [1981]). In the traditional mixed logit literature one can view the individual n as deriving random utility Δ from alternative j in task t, given by

$$\Delta_{njt} = \beta_n x_{njt} + \varepsilon_{njt} \tag{1}$$

where β_n is a vector of coefficients specific to subject n, x_{njt} is a vector of observed attributes of individual j and/or alternative j in task t, and ε_{njt} is a random term that is assumed to be identically and independently distributed extreme value. We use the symbol Δ for utility in (1), since we will need to generalize to allow for non-linear utility functions, and expected utility functionals, and prefer to think of (1) as defining a latent index rather than as utility. In our experience, this purely semantic difference avoids some confusions about interpretation.

Specifically, for our purposes we need to extend (1) to allow for non-linear functions G defined over β and the values of x, such as

$$\Delta_{njt} = G(\beta_n, x_{njt}) + \varepsilon_{njt} \tag{2}$$

For example, x might consist of the vector of monetary prizes m_k and probabilities p_k, for outcome k of K in a given lottery, and we might assume a Constant Relative Risk Aversion (CRRA) utility function

$$U(m_k) = m_k^{\,r} \tag{3}$$

where r is a parameter to be estimated.[3] Under expected utility theory (EUT) the probabilities for each outcome are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery j:

$$EU_j = \Sigma_k \, [\, p_k \times U(m_k) \,] \tag{4}$$

If we let β=r here, we will want to let $G(\beta_n, x_{njt})$ be defined as

---

[2] It is trivial to allow J and T to vary with the individual, but for ease of notation we omit that generality.
[3] The choice of the power function is purely for pedagogical reasons and to keep the exposition simple.

$$G(r_n, m_{njt}, p_{njt}) = EU_j \tag{5}$$

using (3) and (4), and hence let the latent index Δ in (2) be evaluated.[4]

The extension from a linear mixed logit specification, assuming (1), to a non-linear mixed logit specification, assuming (2), has an attractive side-benefit when it comes to identifying the effects of demographic variables such as sex. In the usual specification with linear latent indices of utility the effects of attribute-invariant effects drop out, and one can only consider them by considering interactions with attributes. In effect, the non-linearity of (2) builds this interaction in at a structural theoretical level. An appendix explains this point more formally, although it is probably perfectly intuitive.

The intuition derives from the fact that only differences in (expected) utility matter for choice. Thus one can re-normalize (expected) utility more or less at will, with minor mathematical constraints, as long as the (expected) utility numbers have the same ordering. Then, as Train [2003; p.25] notes:

> The same issue affects the way that socio-demographic variables enter a model. Attributes of the alternatives, such as the time and cost of travel on different modes, generally vary over alternatives. However, attributes of the decision maker do not vary over alternatives. They can only enter the model if they are specified in ways that create differences in utility over alternatives.

If the sex of the agent affects the risk attitude, then this characteristic will affect the (expected) utility evaluation of a given lottery, since each lottery will typically have attributes given by probabilities and outcomes that vary across the two alternatives presented to the subject in any given choice. In effect, the non-linear specification (2) naturally builds in the effect that characteristics have on utility differences.

Returning to the general notation, the population density for β is denoted $f(\beta \mid \theta)$, where θ is a vector defining what we refer to as the hyper-parameters of the distribution of β. Thus individual realizations of β, such as β_n, are distributed according to some density function f. For example, if f

---

[4] This approach generalizes immediately to non-EUT models in which there are more parameters, say to account for probability weighting and loss aversion. It also generalizes to non-CRRA specifications within EUT models that allow for more flexible specifications of risk attitudes that might vary with the level of the prizes. Each of these extensions involves more non-linearities than our EUT example, taking us even further from the domain of linear mixed logit.

is a Normal density then θ_1 would be the mean of that density and θ_2 the standard deviation of that density, and we would estimate the hyper-parameters θ_1 and θ_2. Or *f* could be a Uniform density and θ_1 would be the lower bound and θ_2 would be the upper bound. If β consisted of more than two parameters, then θ might also include terms representing the covariance of those parameters.

Conditional on β_n, the probability that the subject n chooses alternative i in task t is then given by the conditional logit formula, modestly extended to allow our non-linear index

$$L_{nit}(\beta_n) = \exp\{G(\beta_n, x_{nit})\} \;/\; \Sigma_j \exp\{G(\beta_n, x_{njt})\} \tag{6}$$

The probability of the observed choices by subject n, over all tasks T, again conditional on knowing β_n, is given by

$$P_n(\beta_n) = \Pi_t \; L_{ni(n,t)t}(\beta_n) \tag{7}$$

where i(n,t) denotes the lottery chosen by subject n in task t, following the notation of Revelt and Train [1998]. The unconditional probability involves integrating over the distribution of β:

$$P_n(\theta) = \int P_n(\beta_n)\, f(\beta \,|\, \theta)\; d\beta \tag{8}$$

and is therefore the weighted average of a product of logit formulas evaluated at different values of β, with the weights given by the density *f*.

We can then define the log-likelihood by

$$LL(\theta) = \Sigma_n \ln P_n(\theta) \tag{9}$$

and approximate it numerically using simulation methods, since it cannot be solved analytically. Using the methods of Maximum Simulated Likelihood (MSL) reviewed in Train [2003; §6.6, ch.10] and Cameron and Trivedi [2005; ch.12], we define the simulated log-likelihood by taking h=1,...,H replications β^h from the density $f(\beta \,|\, \theta)$:

$$SLL(\theta) = \Sigma_n \ln \{ \Sigma_h P_n(\beta^h)/H \} \tag{10}$$

The core insight of MSL is to evaluate the likelihood conditional on a randomly drawn β^h, do that H times, and then simply take the unweighted average over all H likelihoods so evaluated. The average is unweighted since each replication h is equally likely, by design. If H is "large enough," then MSL converges, under modest assumptions, to the Maximum Likelihood (ML) estimator.[5]

The value of this extension to non-linear mixed logit might not be obvious, because of widespread reliance on theorems showing that the linear mixed logit specification can approximate arbitrarily well any random-utility model (McFadden and Train [2000]; Train [2003; §6.5]).[6] So, why does one need a non-linear mixed logit specification? The reason is that these results only go in one direction: for any specification of a latent structure, defined over "deep parameters" such as risk preferences, they show that there exists an equivalent linear mixed logit. But they do not allow the direct recovery of those deep parameters in the estimates from the linear mixed logit. The deep parameters, which may be the things of interest, are buried in the estimates from the mixed logit, but can only be identified with restrictive assumptions about functional form.[7] For example, risk attitudes can be considered using a linear specification if one assumes that utility is quadratic or that the distribution of returns are Normal (e.g., Luenberger [1998; §9.5]); neither are palatable assumptions in general.

Our specification has been couched in the language of estimating the structural parameters of a model of risk attitudes, but is perfectly general. Another obvious example would be the use of

---

[5] An important practical consideration with MSL is the manner in which replicates are drawn, and the size of H that is practically needed. We employ Halton draws to provide better coverage of the density than typical uniform number generators: see Train [2003; ch.9] for an exposition, and Drukker and Gates [2006] for the numerical implementation we employ. All results use H=250, which is generally large in relation to the literature. Our computational implementation generalizes the linear mixed logit program developed for *Stata* by Hole [2007].

[6] There are reasons to be suspicious of these theorems, although that is not critical for the point being made here. Specifically, two critical assumptions seem to connect observables and unobservables in a highly restrictive way. In one case, the correct claim is made (p.449) that a "primitive postulate of preference theory is that tastes are established prior to assignment of resource allocations." But this does not justify the assumption that "consumers with similar observed characteristics will have similar distributions of unobserved characteristics." Then a related, second assumption is made about attributes. Here the correct claim is that another "primitive postulate of consumer theory is that the description of a resource allocation does not depend on consumer characteristics. Thus, consumers' tastes and perceptions do not enter the 'objective' description of a resource allocation, although they will obviously enter the consumer's evaluation of the allocation." But it does not follow from this observation that "discrete alternatives that are similar in their observed attributes will have similar distributions of unobserved attributes." These assumptions are akin to the identifying assumptions of "random effects" specifications, that the random effect is orthogonal to the observed characteristics used as regressors. One other concern with these theorems is that they rest on polynomial approximations to random utility (McFadden and Train [2000; p. 466]), and these are known to have unreliable properties in statistical applications (e.g., White [1980; §2]). Referring to the class of approximations, including the polynomial, that are generated by applications of Taylor's Theorem, Gallant [1981; p. 212] notes that this "... theorem fails rather miserably as a means of understanding the statistical behavior of parameter estimates and test statistics."

[7] To take one example, of some importance for stated choice models of recreation demand, Herriges and Phaneuf [2002] elegantly show how one can "trick" a linear mixed logit specification into allowing for nested utility structures. They still, however, assume that the indirect utility from alternatives within each nest is linear in attributes and income (p.1086), consistent with a restrictive Cobb-Douglas utility specification.

technology or transportation choices to recover the structural parameters of production functions, or the use of stated or revealed choices to recover the structural parameters of utility functions defined over consumption goods. The analyst needs to fill in their own equations for our (3) and (4), but only need in the end to define $G(\beta_n, x_{njt})$ in (5) and hence in (2).

2. Flexible Population Distributions "For Free"

In principle the mixed logit specification, whether linear or non-linear, allows a wide range of shapes for the probability distribution used to characterize the population. In practice, one typically sees a relatively simple set of distributions used: univariate or multivariate Normal distributions, log-Normal distributions for coefficients known to be non-negative, uniform distributions, or triangular distributions.

One attractive option, since we are already allowing non-linear transformations of the population parameters, is to employ a transformation of the Normal distribution known as the Logit-Normal (L-N) distribution. Originally proposed by Johnson [1949] as an excellent, tractable approximation to the Beta distribution, it has been examined by Mead [1965], Aitchison and Begg [1976; p.3], and Lesaffre, Rizopoulos and Tsonaka [2007]. One nice property of the L-N distribution is that MSL algorithms developed for univariate or multivariate Normal distributions can be applied directly, providing one allows non-linear transformations of the structural parameters, and that is exactly what we are doing already to estimate structural parameters.

Figures 1 and 2 illustrate the wide array of distributional forms that are accommodated by the L-N distribution. The bi-modal and skewed distributions that are possible are particularly attractive. Note that these alternatives are all generated by different values of the two parameters of the (univariate) Normal distribution, so there is no "extra cost" of this flexibility in terms of additional parameters.

One limitation, of course, is that the Beta distribution and the L-N approximation of it, are defined over the unit interval. For some important inferential purposes, such as estimating a subjective probability, this is not a concern, but in general we would like something that is more general. In many other cases though, one would want the estimated distribution to be constrained to lie within specific boundaries dictated by theory. Examples include non-negativity constraints to ensure monotonicity and non-satiation in utility, or restrictions to the unit interval for probabilities or shares. In fact, the power utility function (3) that we employ here for illustration requires that $r>0$ to ensure monotonicity. It is a simple matter to define the so-called "Beta4 distribution" with two additional parameters: one to stretch out the distribution or squeeze it up, and another parameter to shift it left or right. This flexibility makes it possible to theoretically constrain the distribution of the structural parameter to be estimated.[8]

3. An Application: Estimating Risk Attitudes

A. Data

The data consists of N=63 subjects making T=60 choices over J=2 lottery pairs, and where each lottery consisted of up to K=4 outcomes defined over non-negative monetary prizes. The subject was presented a "pie display" of each lottery, and asked to pick one. Outcomes were $0, $5, $10 and $15, and probabilities were given to subjects in ⅛ increments. We ignore expressions of indifference, which were rare.[9] Each subject was given a series of practice choices, which consisted of 4 choices that were played out exactly as the paid choices would be, although not for any earnings. Subjects were paid for 3 of the 60 choices, selected at random after all choices were made.[10]

[8] Obviously some constraints can be accommodated by well-known transformations, such as non-negativity and the Log-normal. This alternative is often standard in linear mixed logit specifications (e.g., Hole [2007; p.390] and the "ln" option). Our approach is more general, particularly for estimates constrained *a priori* to some finite interval.

[9] The extension needed to accommodate the choice of indifference in the model is trivial. It is just a third option in the logit model. We keep it out for the sake of simplicity.

[10] Harrison and Rutström [2008; §3.8] demonstrate that the payment of three lotteries instead of one makes no difference for estimated risk attitudes using this task and a sample drawn from the same population.

B. The Specific Model

Assume a simple CRRA specification under EUT, given by (3) and (4). The parameter r defines the risk attitudes of the subject: $0<r<1$ implies risk aversion, $r=1$ implies risk neutrality, $r>1$ implies risk loving, and $r<0$ implies that subjects violate the assumption of non-satiation. If we ignore observable covariates in the traditional ML estimation approach, this implies that we are assuming homogenous preferences across subjects: the sampling error on r may, of course, reflect the fact that risk preferences can vary across individuals, but the interpretation of the latent process is as if there is one "representative individual" making choices across the sample.

It is common in the analysis of choice data to allow for subjects to make behavioral errors. In effect this is a story about the latent choice process, although there are obvious, and not-so-obvious, econometric implications (e.g., see Harrison and Rutström [2008] and Wilcox [2008] for reviews). We illustrate the effect of this popular extension with the Fechner specification employed by Hey and Orme [1994] and others. In this case there is a new parameter $\mu>0$ that is used to scale the difference in EU of the two lotteries up or down, before one then applies a "link function" to generate a predicted probability of choosing one or the other lottery. In our case the link function is the logistic, but the Fechner correction works before that link function is applied, to transform the latent difference in EU that the logistic function evaluates. If $\mu=1$ then there is no behavioral error; if $\mu<1$ then the given difference in EU is increased, and the choice that was predicted when $\mu=1$ is now even more likely; conversely, if $\mu>1$ then the given difference in EU is decreased, and the choice that was predicted when $\mu=1$ is now less likely to occur. Indeed, as μ gets larger and larger, the predicted choice converges to indifference. This extension amounts to a change in the specification that is captured in the non-linear $G(\cdot)$ function, so that we would have

$$G(r_n, m_{njt}, p_{njt}) = EU_j/\mu \qquad (5')$$

instead of (5). The Fechner error parameter μ is now a part of β_n along with the deep parameter r. We can treat the Fechner parameter as a random or non-random covariate, of course.

C. Estimates [11]

Table 1 contains all estimates for the structural risk aversion parameter r of interest here. We do not list estimates of ancillary parameters.

The MSL estimation approach allows for the parameter r to be normally distributed, to reflect possible heterogeneity across subjects. In this instance there will still be sampling errors on the hyper-parameters characterizing that normal distribution, the mean and standard deviation of the population parameter r. But the possibility that the estimated standard deviation of the population parameter r is positive, and statistically significantly positive, is what differentiates the MSL approach from the ML approach. We initially assume that the parameter is normally distributed, and then consider the extension to the more flexible Beta4 distribution.

If we start by constraining the standard deviation of the population parameter to zero, and treat the Fechner parameter μ as non-random, we literally replicate in Panel I of Table 1 the estimates obtained with ordinary ML: the point estimate of the population mean r is 0.47, with a standard error of 0.014 and a 95% confidence interval between 0.44 and 0.50. For future reference, Panel II extends this specification to allow for one demographic covariate, whether the subject was female, to affect the structural parameter r in a linear fashion. Hence we see a common, if not universal, result: females are more risk averse than men, and the effect is statistically significant.

Now consider the MSL estimates. Allowing the standard deviation of the population parameter r to be non-zero, in Panel III we estimate the mean of r to be 0.38 with a standard error of 0.072, and we estimate the standard deviation of r to be 0.38 with a standard error of 0.06. In Panel III we allow for unobserved heterogeneity, but do not allow for the observed heterogeneity of the sex of the subject. Panel IV undertakes that extension, illustrating one of the side-benefits of an explicitly non-linear mixed logit specification: the ability to study the effects of demographics without having to interact them with choice characteristics. In this case we have an effect of the sex of the subject on the estimated mean of the population distribution of r as well as an effect of sex on

[11] In an appendix, available on request, we document general software to implement the non-linear mixed logit model in *Stata*.

the estimated standard deviation of the population distribution of r. Females are more risk averse on average, but there is much more variability in the population of risk attitudes for females.

The log-likelihood of the specification in Panel III is better than for the specification in Panel II, suggesting in this case that it is more important to control for unobserved heterogeneity than the single observed characteristic we consider here. This need not be a general result. It is quite possible that a longer list of observed characteristics could generate a log-likelihood than just allowing this individual heterogeneity to be handled by the random coefficients approach.[12] In general we want to use observed characteristics as well as allow for unobserved heterogeneity.

Allowing μ to be random along with r, in Panel V we estimate the mean of r to be 0.43 with a standard error of 0.037, and we estimate the standard deviation of r to be 0.329 with a standard error of 0.03. Allowing heterogeneity in the population Fechner parameter slightly improves the precision of the estimates of the core structural parameter r, as one would expect. Figure 3 displays the estimated population distribution corresponding to these estimates, and suggests that there is considerable heterogeneity in the distribution of risk attitudes in this population. A normal assumption in decision theory is to assume non-satiation. Figure 3 shows that the bulk of the distribution of r is estimated to be in the positive domain, but a significant part of the distribution is estimated to lie in the negative domain, violating local non-satiation. This is the curse of assuming a Normal distribution of the structural parameter, which might make theoretical sense at the mean, but clearly implies nonsense in the tails.[13]

Finally, Panel VI considers the extension to the Beta4 distribution.[14] In this case the estimated distribution tightens considerably. The estimates in Panel VI of Table 1 do not have much

[12] As it happens, adding our "standard list" of demographic covariates, reflecting ethnicity, major and GPA level, improves the log-likelihood of the ML specification significantly, but only results in a log-likelihood of -2382.9.

[13] Although we are treating non-satiation as a given, it is apparent that one could use our approach to generate nested hypothesis tests of that assumption.

[14] We estimate the two parameters that "shift" and "stretch" the parameters of the logit-normal distribution. The first is estimated to be -0.094 and the second is estimated to be 2.40. In our experience these values can easily be set parametrically based on inspection of the implied population distribution, or they can be set based on what theory tells us they should be. These parametric bounds facilitate rapid estimation, and provide excellent starting values for the final estimation stage reported here. As it happens, the lower bound for r, -0.094, is not statistically significantly different from 0 (*p*-value = 0.75).

intuition by themselves, and can be best appreciated by inspecting Figure 4, where we also show the implied Beta4 distribution overlaid on the Normal distribution from Figure 3 for comparison. The lower tail of the Beta4 distribution is sharply truncated at 0, thus telling us that only a minority of the choices show a violation of non-satiation. Formal tests of the Beta4 distribution, following D'Agostino, Belanger, and D'Agostino [1990], lead us to reject the hypothesis that it is a Normal distribution at any standard level of significance.

4. Comparison to the Literature

There is a large econometric literature on "distribution free" estimators of the binary choice model, reviewed well in Pagan and Ullah [1999; ch.7]. Using our random utility interpretation, the term distribution-free in this case refers the relaxation of parametric assumptions about the form of the random error terms ε_{njt} of (1). As is well-know, and reviewed historically by McFadden [2001], making parametric assumptions about these error terms, such as that they are distributed as type I extreme (or Gumbel), is the same as assuming directly that the latent index Δ_{njt} is linked to the choice probabilities by a logit function. The popular alternative parametric assumption to the logit is of course the cumulative normal distribution. The econometric literature on distribution free estimators relaxes these parametric assumptions. That is important, but is not the same thing as relaxing the assumption that the deterministic part of the latent index, $G(\beta_n, x_{njt})$ in our (2), be linear. This point is made explicitly by Pagan and Ullah [1999; p. 279], for example.

The econometric literature that relaxes the assumption that $G(\beta_n, x_{njt})$ be linear has two strands. One allows $G(\beta_n, x_{njt})$ to be characterized with a flexible functional form, and is due to Gallant [1981], although there have were no applications of this general approach to the binary choice context until Chen and Randall [1997]. The other strand allows $G(\beta_n, x_{njt})$ to be characterized non-parametrically as a monotone and concave function, and is due to Matzkin [1991].[15] Although it

[15] Matzkin [1992] extends the non-parametric specification of $G(\beta_n, x_{njt})$ to also allow for a distribution free specification of the error terms. This extension raises delicate issues of identification, which of course are far less severe when one makes parametric assumptions about $G(\beta_n, x_{njt})$ and/or the error terms.

is possible in principle to use either of these approaches to approximate non-linear utility functions, it is easier to write those functions out explicitly and estimate the structural parameters directly *if that is what one wants to make inferences about.*

The tradition in theoretical econometrics is to avoid any such assumption about structure. There are some good reasons for this, but also some disadvantages. In this instance, for example, Matzkin [1992; p. 240] argues[16] that

> All these previously developed methods assume that [the systematic component $G(\beta_n, x_{njt})$ in our specification] is known up to a finite dimensional parameter vector. These assumptions are, however, almost never justified. Economic theory does not impose any restrictions on the parametric structure of functions.

The final sentence is of course correct as a general matter, but one should not therefore assume that economists are unwilling to write out explicit functional forms for certain purposes. So we take issue with the phrase "almost never justified," and can point to a myriad of literature in which such parametric assumptions are employed as a standard and uncontroversial matter. This is not to say that those parametric assumptions should remain untested, just to note that there are inferential objectives for which one is willing to accept certain parametric assumptions. Estimating risk attitudes, whether one uses EUT or non-EUT models of decision making, is one in which such parametric assumptions are commonly used, and where the parametric forms *have to be non-linear* to draw any useful substantive inferences.

Thus our approach is to swim against the tide of the econometric literature that wants to relax linearity of $G(\beta_n, x_{njt})$ by being as agnostic about functional form as possible.[17] One reason is the practical inability of these agnostic approaches to deliver estimates of the structural parameters of interest in a tractable computational procedure. When those agnostic approaches are able to deliver those estimates we will be the first to swim with the tide instead of against it.

---

[16] In context this passage also refers to parametric assumptions about the error terms, but we focus on the point about the systematic functional form inside the latent index.

[17] The agnostic approach certainly respects certain fundamental economic properties of the systematic component of the random utility function. Matzkin [1991], for example, maintains monotonicity and concavity. Thus the agnostic approach is not completely atheistic.

5. Conclusions

We develop a non-linear mixed logit specification that allows the analyst to directly estimate the "deep structural parameters" of interest in many behavioral analyses. We illustrated by considering the familiar case of estimating risk preferences from a sample of lottery choices. Since risk attitudes fundamentally entails modeling non-linearities in most behavioral theories, whether the non-linearity is in the utility function and/or a probability weighting function, this application is canonical. The data we consider are generated in controlled laboratory experiments, to ensure that we know all variables of interest to illustrate the approach in a clean manner, but the broader usefulness of this approach is obvious.

Figure 1: Illustrative Symmetric Logit-Normal Distributions

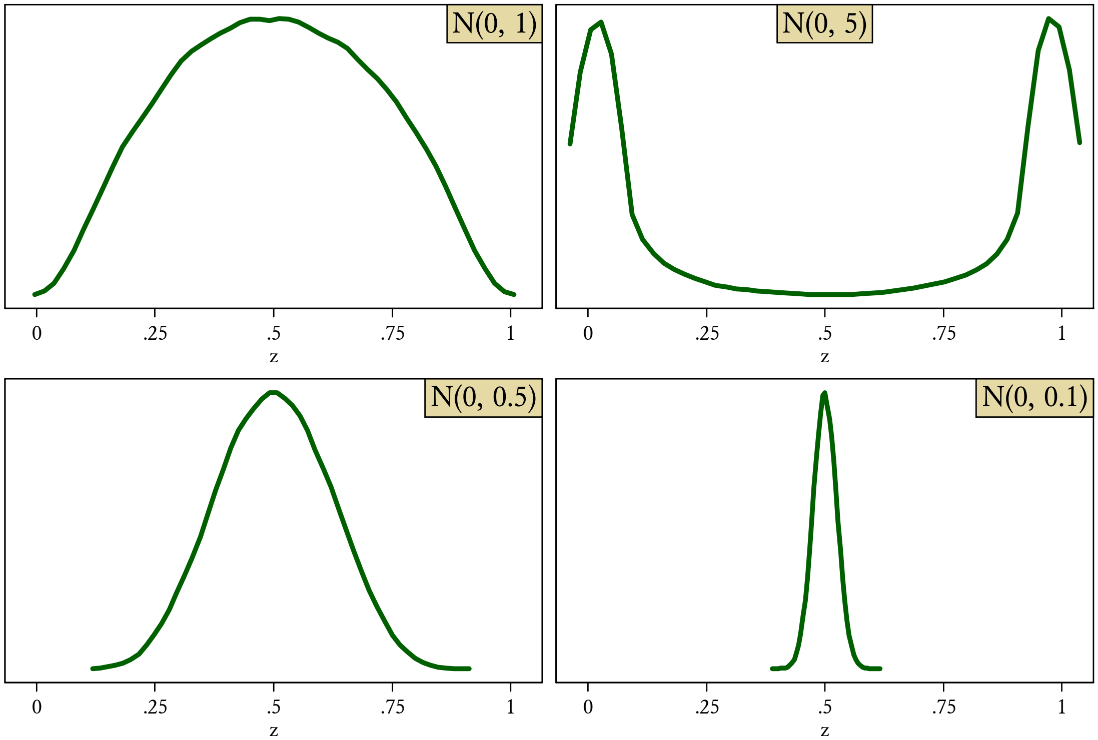


Figure 2: Illustrative Asymmetric Logit-Normal Distributions

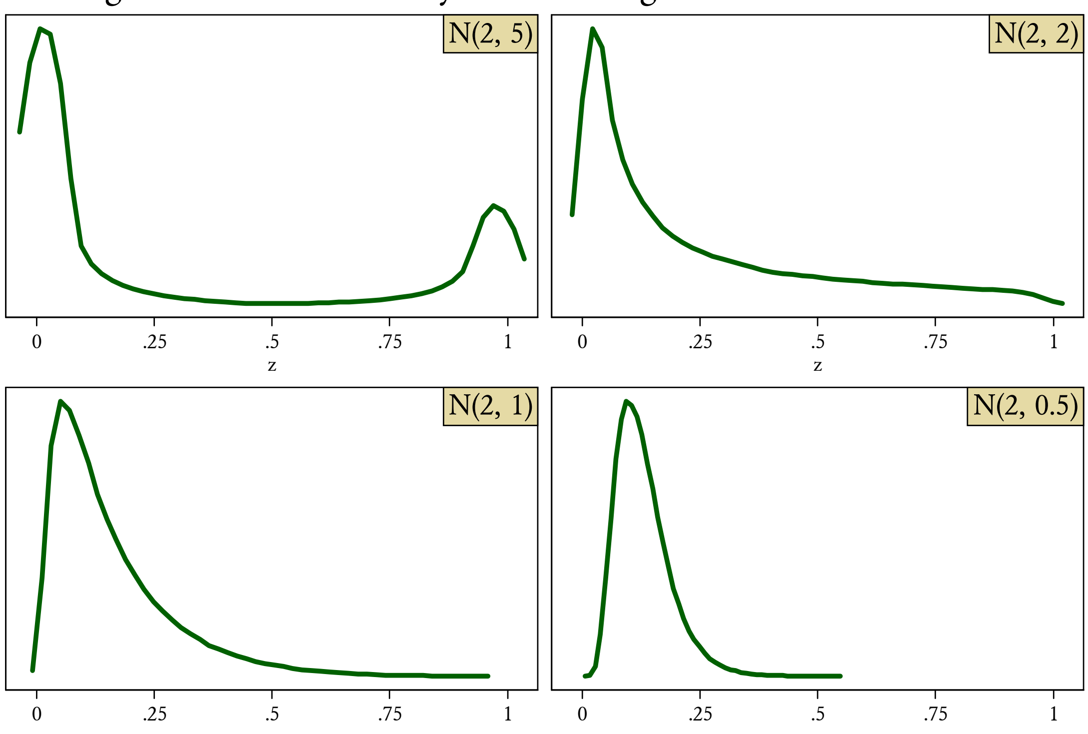


Table 1: Estimates of Structural Risk Aversion Parameter

Estimates of Fechner Parameter μ, and other ancillary parameters, not shown

Parameter	Point Estimate	Standard Error	*p*-value	95% Confidence Interval	
I. ML Estimates with No Heterogeneity, Constrained Normal (LL= -2415.8)					
r_{mean}	0.468	0.014	<0.001	0.439	0.496
r_{sd}	0 (constrained)	0 (constrained)			
II. ML Estimates with Observable Heterogeneity, Constrained Normal (LL= -2397.9)					
r_{mean} for all	0.545	0.016	<0.001	0.514	0.578
r_{mean} for females	-0.170	0.031	<0.001	-0.231	-0.108
r_{sd}	0 (constrained)	0 (constrained)			
III. MSL Estimates with Unobserved Heterogeneity, Non-Random μ, Normal (LL= -2193.5)					
r_{mean}	0.383	0.072	<0.001	0.241	0.524
r_{sd}	0.378	0.062	<0.001	0.256	0.500
IV. MSL Estimates with Observed & Unobserved Heterogeneity, Non-Random μ, Normal (LL= -2184.6)					
r_{mean} for all	0.515	0.050	<0.001	0.417	0.613
r_{mean} for females	-0.354	0.096	<0.001	-0.542	-0.167
r_{sd} for all	0.264	0.038	<0.001	0.190	0.338
r_{sd} for females	0.403	0.088	<0.001	0.238	0.570
V. MSL Estimates with Unobserved Heterogeneity, Random μ, Bivariate Normal (LL= -2120.8)					
r_{mean}	0.430	0.037	<0.001	0.358	0.502
r_{sd}	0.329	0.025	<0.001	0.279	0.378
VI. MSL Estimates with Unobserved Heterogeneity, Random μ, Bivariate Beta4 (LL= -2119.3)					
r_{mean}	1.587	1.503	0.291	-0.242	2.934
r_{sd}	0.493	0.578	0.162	-1.903	0.317

Figure 3: Population Distribution for CRRA Parameter
Assuming a Normal Distribution

CRRA=1 is risk neutral, CRRA <1 is risk averse, and CRRA>1 is risk loving.
Maximum likelihood estimate is .47.

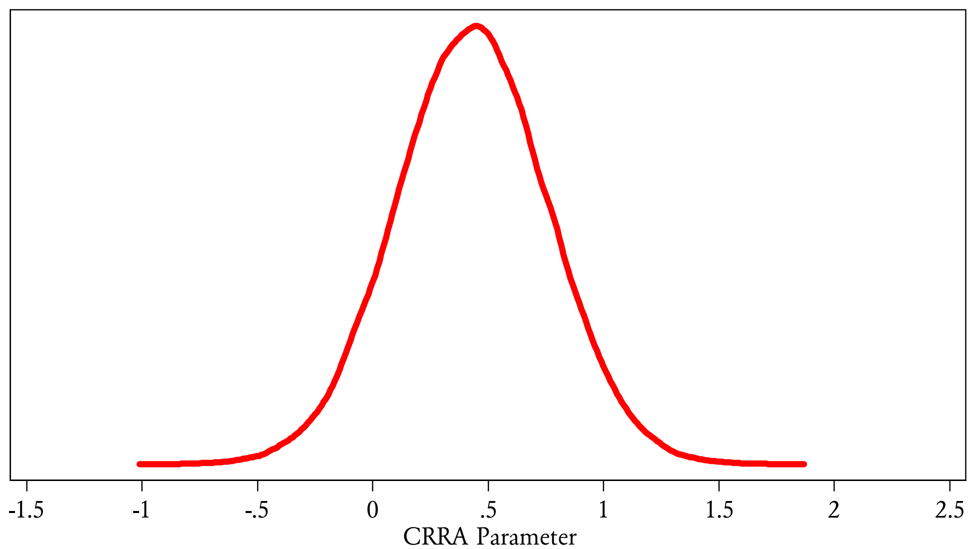
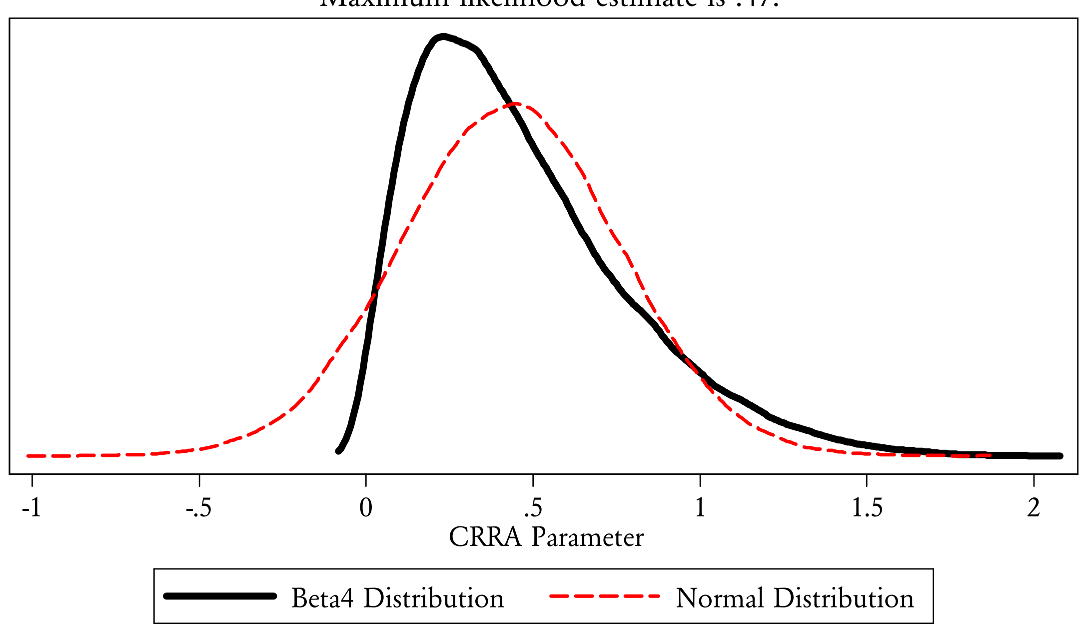


Figure 4: Population Distribution for CRRA Parameter
Assuming Beta4 or Normal Distributions

CRRA=1 is risk neutral, CRRA <1 is risk averse, and CRRA>1 is risk loving.
Maximum likelihood estimate is .47.



Appendix: Identification of Demographic Effects

It is well known that choice-invariant characteristics drop out of the usual linear specifications, and that they can only be included if they are interacted with choice-varying attributes. The use of non-linear latent indices for utility builds this sort of "interaction" in automatically, and in a theoretically natural manner. Although this may be obvious on inspection of (1) and (2), it is perhaps worth spelling out more formally. We use the exposition from Greene [2008; §0.3], although the point is well known (e.g., Train [2003; §2.5.1]). Recall the notation in the text, where individual n is viewed by the econometrician as deriving utility Δ from alternative j in task t, given by

$$\Delta_{njt} = \beta_n x_{njt} + \varepsilon_{njt} \tag{A1}$$

where β_n is a vector of coefficients specific to subject n, x_{njt} is a vector of observed attributes of individual j and/or alternative j in task t, and ε_{njt} is a random term. Focus on just one task, so we can drop the subscript t. Assume that the vector x is broken up into two components: a vector w that varies with the alternative j and (possibly) with the individual n, and a vector z that does not vary with the alternative j, but does differ across individuals. In our lottery choice setting, elements of w include the probabilities and outcomes of lottery j, and elements of z might include the sex of the individual.

We can then assume that the individual evaluates utility using these linear specifications, for j $\in$ {A, B}:

$$\Delta_{nA} = \eta_n w_{nA} + \gamma_n z_n + \varepsilon_{nA} \tag{A2}$$
$$\Delta_{nB} = \eta_n w_{nB} + \gamma_n z_n + \varepsilon_{nB} \tag{A3}$$

If we observe choice A then this choice reveals that

$$\Delta_{nA} > \Delta_{nB} \tag{A4}$$

Hence we have

$$\eta_n w_{nA} + \gamma_n z_n + \varepsilon_{nA} > \eta_n w_{nB} + \gamma_n z_n + \varepsilon_{nB} \tag{A5}$$

or

$$(\eta_n w_{nA} - \eta_n w_{nB}) + (\gamma_n z_n - \gamma_n z_n) > \varepsilon_{nB} - \varepsilon_{nA} \tag{A6}$$

and therefore

$$(\eta_n w_{nA} - \eta_n w_{nB}) > \varepsilon_{nB} - \varepsilon_{nA} \tag{A7}$$

So the vector z has dropped out completely.

When we replace the linear latent index (A1) with the non-linear index,

$$\Delta_{njt} = G(\beta_n, x_{njt}) + \varepsilon_{njt} \tag{A1$'$}$$

we have counterparts to (A2) and (A3) as

$$\Delta_{nA} = G(\eta_n, \gamma_n, w_{nA}, z_n) + \varepsilon_{nA} \tag{A2$'$}$$
$$\Delta_{nB} = G(\eta_n, \gamma_n, w_{nB}, z_n) + \varepsilon_{nB} \tag{A3$'$}$$

Thus choice observation (A4) only implies

$$G(\eta_n, \gamma_n, w_{nA}, z_n) + \varepsilon_{nA} > G(\eta_n, \gamma_n, w_{nB}, z_n) + \varepsilon_{nB} \tag{A5$'$}$$

as a counterpart to (A5), and

$$G(\eta_n, \gamma_n, w_{nA}, z_n) - G(\eta_n, \gamma_n, w_{nB}, z_n) > \varepsilon_{nB} - \varepsilon_{nA} \tag{A6$'$}$$

as a counterpart to (A6), and the vector z simply does not drop out without more structure being placed on $G(\cdot)$. For the applications we consider, and that would be natural for virtually all models of risk attitudes, no further simplification beyond (A6$'$) is possible.

References

Aitchison, J., and Begg, C.B., "Statistical Diagnosis when Basic Cases are Not Classified with Certainty," *Biometrika*, 63(1), 1976, 1-12.

Binswanger, Hans P., "Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India," *Economic Journal*, 91, December 1981, 867-890.

Cameron, A. C., and Trevedi, K.P., *Microeconometrics: Methods and Applications* (New York: Cambridge University Press, 2005).

Chen, Heng Z., and Randall, Alan, "Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation," *Journal of Econometrics*, 76, 1997, 323-340.

D'Agostino, R. B., Belanger, A.J., and D'Agostino, R.B. Jr., "A Suggestion for using Powerful and Informative Tests of Normality," *American Statistician*, 44, 1990, 316-321.

Drukker, David M., and Gates, Richard, "Generating Halton sequences using Mata," *Stata Journal*, 6(2), 2006, 214-228.

Gallant, A. R., "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form," *Journal of Econometrics*, 15, 1981, 211-245.

Greene, William, "Discrete Choice Modelling," in T.C. Mills and K. Patterson (eds.), *Palgrave Handbook of Econometrics. Volume 2: Applied Econometrics* (London: Palgrave, 2008).

Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).

Herriges, Joseph A., and Phaneuf, Daniel J., "Inducing Patterns of Correlation and Substitution in Repeated Logit Models of Recreation Demand," *American Journal of Agricultural Economics*, 84(4), November 2002, 1076-1090.

Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.

Hole, Arne Risa, "Fitting Mixed Logit Models by Using Maximum Simulated Likelihood," *Stata Journal*, 7(3), 2007, 388-401.

Johnson, N.L., "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, January 1949, 149-176.

Lesaffre, Emmanuel; Rizopoulos, Dimitris, and Tsonaka, Roula, "The Logistic Transform for Bounded Outcome Scores," *Biostatistics*, 8(1), 2007, 72-85.

Luenberger, David G., *Investment Science* (New York: Oxford University Press, 1998).

Matzkin, Rosa L., "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, 59(5), 1991, 1315-1327.

Matzkin, Rosa L., "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold Crossing Models," *Econometrica*, 60(2), 1992, 239-270.

McFadden, Daniel, "Economic Choices," *American Economic Review*, 91, 2001, 351-378.

McFadden, Daniel, and Train, Kenneth, "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 2000, 447-470.

Mead, R., "A Generalised Logit-Normal Distribution," *Biometrics*, 21(3), September 1965, 721-732.

Pagan, Adrian, and Ullah, Aman, *Nonparametric Econometrics* (New York: Cambridge University Press, 1999).

Revelt, David, and Train, Kenneth, "Mixed Logit With Repeated Choices: Households' Choices of Appliance Efficiency Levels," *Review of Economics & Statistics*, 80, 1998, 647-657.

Train, Kenneth E., *Discrete Choice Methods with Simulation* (New York: Cambridge University Press, 2003).

White, Halbert, "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21, 1980, 149-170.

Wilcox, Nathaniel T., "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison," in J. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008).