

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

**Working Paper**

## Non-Linear Mixed Logit and the Characterization of Individual Heterogeneity

Working paper, No. 6-2009

**Provided in Cooperation with:**

Department of Economics, Copenhagen Business School (CBS)

*Suggested Citation:* Andersen, Steffen; Harrison, Glenn W.; Hole, Arne Risa; Rutström, Elisabet E. (2009) : Non-Linear Mixed Logit and the Characterization of Individual Heterogeneity, Working paper, No. 6-2009, Copenhagen Business School (CBS), Department of Economics, Frederiksberg, <https://hdl.handle.net/10398/7802>

This Version is available at:

<https://hdl.handle.net/10419/208559>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-nd/3.0/>



**Copenhagen  
Business School**  
HANDELSHØJSKOLEN

# **Department of Economics**

Copenhagen Business School

**Working paper 6-2009**

## **NON-LINEAR MIXED LOGIT AND THE CHARACTERIZATION OF INDIVIDUAL HETEROGENEITY**

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

# Non-Linear Mixed Logit and the Characterization of Individual Heterogeneity

by

Steffen Andersen, Glenn W. Harrison, Arne Risa Hole and E. Elisabet Rutström<sup>†</sup>

February 2009

*Working Paper 09-02*, Department of Economics,  
College of Business Administration, University of Central Florida, 2007

*Abstract.* Experimental data exhibit considerable individual heterogeneity. We review the econometric methods employed to characterize that heterogeneity. We pay particular attention to the trade-off between collecting and allowing for observable characteristics, such as the familiar demographics, and the use of statistical methods to allow for unobserved individual heterogeneity. We demonstrate that these tools are complementary.

<sup>†</sup> Department of Economics and Centre for Economic and Business Research, Copenhagen Business School, Copenhagen, Denmark (Andersen); Department of Economics, College of Business Administration, University of Central Florida, USA (Harrison and Rutström); and Department of Economics, University of Sheffield, UK (Hole). E-mail contacts: sa.eco@cbs.dk, gharrison@research.bus.ucf.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, and the Danish Social Science Research Council for research support under project #275-08-0289.

The rich data generated by experiments has spurred interest in carefully characterizing the individual heterogeneity of behavioral response. One immediate reason for this interest is to make inferences about differences in behavior. For example, are women more risk averse than men? Another reason is to evaluate comparative static predictions of theory, since randomization to treatment is often insufficient to ensure the absence of confounding differences in samples.

There are three broad approaches to allowing for individual heterogeneity in experimental data. The first is to collect enough observations from each subject to be able to estimate models for that individual. Hey and Orme [1994] remains a shining exemplar of this approach. Unfortunately, this option is very often infeasible, or prohibitively expensive, and recourse to data pooled over individuals becomes necessary.<sup>1</sup>

The second approach is to collect a standard list of observable demographic characteristics from the individual, pool the responses across all subjects, and condition estimation of parameters on those characteristics. Thus one might estimate a model in which a risk aversion parameter is a linear function of sex, age, race, income, and so forth. In this case one obtains estimates of the coefficients of the effect of each of the characteristic on the parameter, as well as a constant term (e.g., Harrison, Lau and Rutström [2007]). The danger with this approach is that there may be some characteristic of individuals that is not observed by the experimenter<sup>2</sup> and that systematically affects responses: this is the so-called “unobserved individual heterogeneity” problem.<sup>3</sup>

The third approach to allowing for individual heterogeneity is to use econometrics methods that are relatively agnostic about the form that heterogeneity takes, but that allow explicitly for it. This approach might be usefully termed the “random coefficients” approach. Instead of assuming that a risk aversion coefficient was a linear function of observables, one might assume that the

---

<sup>1</sup> One direction for future research would be to take data that allows inferences to be made at the level of the individual and see if one can reliably recover the preferences estimated at the individual level using pooled-data approaches. In effect, this proposes the use of these experimental data as a Monte Carlo test-bed for the quality of the pooled-data estimators.

<sup>2</sup> One occasionally hears comments that one should not bother collecting *any* individual characteristics since we can never know the true set of characteristics, but this position is obviously dogmatic.

<sup>3</sup> We say so-called, since the term “unobserved” is not literally correct. We do observe that a certain set of observations were generated by one subject, and another set by other subjects, and this observation is used in all standard methods to correct for unobserved individual heterogeneity.

coefficient was normally distributed across the sample. In effect, each subject would be assumed to have some true coefficient value, but these values are viewed as being distributed across the sample in a way that can be characterized by a normal distribution. One then estimates *hyper-parameters* to reflect this population distribution.<sup>4</sup>

For example, if one assumes a normal distribution, one would estimate a mean (population) risk aversion coefficient and a standard deviation in the (population) risk aversion coefficient. Each *hyper-parameter* would be estimated by a point estimate and a standard error.<sup>5</sup> As the sample size gets larger, one would expect the estimated standard errors of the *hyper-parameters* to shrink, for the usual reasons, but one would not necessarily expect the standard deviation of the population *parameter* to shrink. This approach generalizes naturally to non-normal distributions for the population parameter, and to multi-variate distributions where there are more than one population parameter.

There is no necessary tension between the second and third approach.<sup>6</sup> However, in practice one tends to see data evaluated using one or the other method. We evaluate the relative role in each approach in a canonical data set of interest to experimental economists, demonstrating how both approaches can be used in a complementary fashion. To keep things focused, we restrict attention to lottery choices over prospects with known probabilities, which has obviously been a foundational domain of research into risk attitudes and the validity of expected utility theory.

In section 1 we briefly review the now-familiar “mixed logit” specification of random utility models, in which the latent index is assumed to be a linear function of observable characteristics of the choice and/or individual. Our exposition is brief, reflecting the availability of excellent

---

<sup>4</sup> The terminology across the econometric literature is not standard, so the expression “hyper-parameter” can have other meanings.

<sup>5</sup> To be pedantic, there are then four estimates: the point estimate and standard error of the mean of the population parameter, and the point estimate and standard error of the standard deviation of the population parameter.

<sup>6</sup> Nor between the first approach and the latter two. It is likely that one cannot obtain reliable estimates for some models for each individual across a sample, due to errant behavior by those individuals, but that the subsequent analysis requires some estimate. In this case one could complement the individual-level estimate and standard error with a predicted point estimate and standard error from one of the latter two approaches. Providing the nature of this stochastic imputation is properly accounted for, the “complete case analysis” can then proceed over the entire sample. This problem will become practically significant as experiments are increasingly used to augment the data collected in large-scale household surveys in the field, not least because there are then likely to be severe limitations on how many questions one can ask in the experimental task.

treatments by Train [2003] and others, and sets the stage for an extension to allow for the latent index to be a non-linear function of characteristics. This extension is needed since virtually all interesting functional forms for utility and probability weighting involve non-linear functions, but virtually all applications of random utility models assume a linear or linearized approximation of the true utility function.<sup>7</sup>

This extension to non-linear mixed logit, while formally modest, is also likely to be of considerable value more generally, since it allows direct estimation of latent structural parameters of virtually any specification. In an appendix we document general software that we have developed within *Stata*, to estimate non-linear mixed logit models.

The value of this extension 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]).<sup>8</sup> 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 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 linear mixed logit, but cannot be identified

---

<sup>7</sup> One also comes across similar linear approximations to production functions in the econometric literature on stochastic frontier models (e.g., Kumbhakar and Lovell [2000; p.72]). In that case the linear specification corresponds exactly to a Cobb-Douglas production function, or an approximation to that function. Such production functions are not, shall we say, close to the frontier of modeling in economics. Non-linear random coefficient specifications, with logit or related specifications, have appeared in some applications using experimental data: for example, Botti, Conte, DiCagno and D’Ippoliti [2008], Cabrales and García-Fontes [2000], Conte, Hey and Moffatt [2007], De Roos and Sarafidis [2006] and Wilcox [2006; §4][2008][2009].

<sup>8</sup> 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.

without a lot of algebra or some unpalatably restrictive assumptions about functional form.<sup>9</sup>

In section 2 we consider data from a replication by Harrison and Rutström [2008] of the classic study of individual choice under uncertainty due to Hey and Orme [1994]. In contrast to the original study, this replication involved the collection of data on individual characteristics of the sample, allowing a test of the relative role of “observables” and “unobservables” in this design.

In summary, we develop a new econometric specification for non-linear mixed logit models, to allow us to estimate latent parameters reflecting the risk preferences of subjects making binary choices over lotteries. We then use that specification to examine the role of observable characteristics and random coefficients in characterizing heterogeneity in latent risk attitudes.

## 1. Linear and Non-Linear Mixed Logit

Assume a sample of  $N$  subjects making choices over  $J$  lotteries in  $T$  experimental tasks.<sup>10</sup> 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], Eckel and Grossman [2002]). In the traditional mixed logit literature one can view the individual  $n$  as deriving utility  $\Delta$  from alternative  $j$  in task  $t$ , given by

$$\Delta_{njt} = \beta_n x_{njt} + \epsilon_{njt} \quad (1)$$

where  $\beta_n$  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  $\epsilon_{njt}$  is a random term that is assumed to be an identically and independently distributed extreme value. We use the symbol  $\Delta$  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.

---

<sup>9</sup> 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.

<sup>10</sup> It is trivial to allow  $J$  and  $T$  to vary with the individual, but for ease of notation we omit that generality.

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

$$\Delta_{njt} = H(\beta_n, x_{njt}) + \varepsilon_{njt} \quad (2)$$

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

$$U(m_k) = m_k^r \quad (3)$$

where  $r$  is a parameter to be estimated. 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  $j$ :

$$EU_j = \sum_k [p_k \times U(m_k)] \quad (4)$$

If we let  $\beta=r$  here, we will want to let  $H(\beta_n, x_{njt})$  be defined by  $H(r_n, m_{njt}, p_{njt})$  using (3) and (4), and hence let the latent index  $\Delta$  in (2) be evaluated. 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.

The extension from a linear mixed logit specification, assuming (1), to a non-linear mixed logit specification, assuming (2), has a singularly 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 characteristics 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  $\beta$  is denoted  $f(\beta | \theta)$ , where  $\theta$  is a vector defining what we refer to as the hyper-parameters of the distribution of  $\beta$ . Thus individual realizations of  $\beta$ , such as  $\beta_n$ , are distributed according to some density function  $f$ . For example, if  $f$  is a Normal density then  $\theta_1$  would be the mean of that density and  $\theta_2$  the standard deviation of that density, and we would estimate the hyper-parameters  $\theta_1$  and  $\theta_2$ . Or  $f$  could be a Uniform density and  $\theta_1$  would be the lower bound and  $\theta_2$  would be the upper bound. If  $\beta$  consisted of more than two parameters, then  $\theta$  might also include terms representing the covariance of those parameters.

Conditional on  $\beta_n$ , 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

$$L_{nit}(\beta_n) = \exp\{H(\beta_n, x_{nit})\} / \sum_j \exp\{H(\beta_n, x_{njt})\} \quad (5)$$

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

$$P_n(\beta_n) = \prod_t L_{ni(n,t)t}(\beta_n) \quad (6)$$

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  $\beta$ :

$$P_n(\theta) = \int P_n(\beta_n) f(\beta | \theta) d\beta \quad (7)$$

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

We can then define the log-likelihood by

$$LL(\theta) = \sum_n \ln P_n(\theta) \quad (8)$$

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  $r=1, R$  replications  $\beta^r$  from the density  $f(\beta | \theta)$ :

$$SLL(\theta) = \sum_n \ln \{ \sum_r P_n(\beta^r) / R \} \quad (9)$$

The core insight of MSL is to evaluate the likelihood conditional on a randomly drawn  $\beta^r$ , do that  $R$  times, and then simply take the unweighted average over all  $R$  likelihoods so evaluated. The average is unweighted since each replication  $r$  is equally likely, by design. If  $R$  is “large enough,” then MSL converges, under modest assumptions, to the Maximum Likelihood (ML) estimator.

An important practical consideration with MSL is the manner in which replicates are drawn, and the size of  $R$  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. As to the size of  $R$ , we illustrate the importance of allowing for relatively large  $R$  in our applications.

Our computational implementation generalizes the linear mixed logit program developed for *Stata* by Hole [2007], and is documented in more detail in an appendix. Our *Stata* program can be used more generally than the applications we consider.

## 2. A Model with No Covariates

### *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  $1/8$  increments. We ignore expressions of

indifference, which were rare. 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.<sup>11</sup>

### *B. Estimates*

To start, assume a simple CRRA specification under EUT, given by (3) and (4). The parameter  $r$  defines the risk attitudes of the subject:  $r < 1$  implies risk aversion,  $r = 1$  implies risk neutrality, and  $r > 1$  implies risk loving. For the moment, assume that there are no 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. The MSL estimation approach allows for the parameter  $r$  to be normally distributed, to reflect possible heterogeneity across subjects. 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. In fact, we can numerically replicate the ML estimates by adopting MSL and constraining the standard deviation to be exactly zero.

Table 1 contains estimates for the basic set of models considered. Panel A of Table 1 displays the estimates obtained using standard ML methods, and panel A.I shows the estimates assuming no allowance for heterogeneity through observable covariates.<sup>12</sup> In this specification the estimate of  $r$  is equal to 0.539, with a standard error of 0.037, implying a 95% confidence interval between 0.466 and 0.611. Thus there is evidence of risk aversion ( $r < 1$ ), and this finding is statistically significant at the usual levels. This estimate is for the pooled data, with no covariates, so

---

<sup>11</sup> 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.

<sup>12</sup> Coefficients that are italicized in Table 1 are non-random.

it is assumed that every subject uses the same utility function for each choice, and that all subjects have the same risk preferences. This general result, of modest risk aversion for stakes of this level and subject pools of this kind, is well-documented in the literature (e.g., see Harrison and Rutström [2008] for a review).

If we allow the population parameter  $r$  to be normally distributed, then we have two hyper-parameters to estimate from the sample data. This is the random coefficient counterpart of the first specification. What is different now is that we allow for individuals to have different risk preferences, and estimate the extent of unobserved heterogeneity in risk preferences. Panel B of Table 1 displays the MSL estimates, and panel B.I shows the counterpart to panel A.I, in the sense that it contains no observable covariates, but that it allows for the parameter  $r$  to be random. The first hyper-parameter is the population mean of  $r$ , which is estimated using MSL to be 0.532 with a standard error of 0.071. The second hyper-parameter is the population standard deviation of  $r$ , which is estimated to be 0.450 with a standard error of 0.049. So each hyper-parameter is precisely estimated, but we infer considerable heterogeneity of risk preferences in the *population*. On average the subject is risk averse, and there is an 85% chance that the typical subject is also risk averse ( $0.85 = \Phi((1-0.532)/0.450)$ ) if we just focus on the *point estimates* of the hyper-parameters. An appendix documents the numerical procedures followed, and discusses checks for numerical reliability of the estimates.

### *C. Interpretation*

It is important to be careful, to the point of semantic nuance, when making inferences from the estimates obtained from random coefficient specifications. The hyper-parameters of the coefficient distributions are measured with uncertainty. These parameters therefore have their own sampling distributions which should not be confused with the distribution of the coefficient itself.<sup>13</sup>

---

<sup>13</sup> Hensher and Green [2003; p.137] make the same point about mixed logit: “After model estimation, there are many results for interpretation. An early warning – parameter estimates typically obtained from a random parameter or error components specification should not be interpreted as stand-alone parameters but must be assessed jointly with other linked parameter estimates.”

To illustrate the potential importance of this distinction, consider the hypothetical example illustrated in Figure 1. Here we have a point estimate for the mean of some population parameter  $\beta$  that is positive, and a point estimate for the standard deviation of the population parameter that is “small,” in the sense that the top panel conveys. The middle panel shows what one might expect to see from a finite number of draws from the population if the standard error on the estimate of the mean was “large” enough, and if the standard error on the estimate of the standard deviation was very small. It can be seen from the figure that the estimated hyper-parameters along with their standard errors represent a “distribution of distributions,” and inferences drawn from the estimation results should factor in this uncertainty rather than treating the point estimates as if they were the true values of the hyper-parameters.

The bottom panel of Figure 1 shows the sampling distribution of the mean parameter of the population distribution, reflecting the fact, from the large standard error for the mean point estimate, that we have virtually no idea if the mean is positive or negative. In effect, the top panel throws away the information from the MSL estimates that we have very imprecise knowledge of the population mean. It is easy to come up with similar examples where the standard errors of the point estimates lead one to make seemingly different inferences than when one relies solely on the point estimates. Again, these inferences are perfectly consistent: they are just inferences about different things.

If we want to make statements about the expected risk aversion of an individual drawn from this population, we use the point estimate of the hyper-parameter for the mean of  $r$  (0.532 in the above example). If we want to make statements about the distribution of *average* risk attitudes in the population, we evaluate the population distribution using the point estimates of the hyper-parameters for the mean and standard deviation of that distribution (0.532 and 0.450 above). This is what we did when we made the claim about there being “an 85% chance that the typical subject is also risk averse.” But if we want to make statements about the distribution of risk attitudes in the population, we also have to take into account the *standard errors* of the estimates of the hyper-

parameters (0.071 and 0.049 above).

The fact that the parameters are measured with uncertainty is not always taken into account when interpreting the results of mixed logit models. For example, Revelt and Train [1998; p.651] conclude that the “mixed logit implies that approximately 9% of the population place a negative coefficient on savings” in a model of household appliance choice, and where savings here refers to the dollars of savings that a high-efficiency appliance provides in comparison to some standard appliance. This inference derives from point estimates (their Table 1, p.651) of the mean and standard deviation on the savings variable that are 3.03 and 2.24, respectively, since  $0.088 = \Phi((0 - 3.03)/2.24)$ . This statement does not take into account the standard errors on the point estimates, which are 0.345 and 0.281, respectively. If we take a simulated draw of 10,000 from this joint normal distribution and calculate the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles of these draws, we find that the 95% confidence interval on the claim that someone in this population has a negative effect from savings is between 2.7% and 16.7% of the population.<sup>14</sup> So the true proportion could in fact be as low as 2.7% and as high as 16.7% within a conventional range of uncertainty. Following convention, we use point estimates to make statements about the average effect, and the simulated population distribution to make statements about the standard errors of that average effect and about confidence intervals.

### 3. Allowing for Covariates

When we extend the homogeneous agents specification with *non-random* coefficients to allow for a sex effect (panel A.II of Table 1), we estimate  $\alpha$  for men to be 0.696 with a standard error of 0.034 and for women to be 0.321 lower than for men, with a standard error on this differential effect of 0.077 and therefore a  $p$ -value of only 0.011. So we estimate the 95% confidence interval of the effect of sex to be between -0.472 and -0.170. Thus we would conclude that there is evidence of a

---

<sup>14</sup> This calculation assumes that the covariance between the mean and standard deviation coefficients is zero, since it is not reported in Revelt and Train [1998]. With a non-zero covariance the confidence interval may be wider or narrower. Our comparable calculations below take into account the covariance.

statistically significant effect of sex on risk aversion, and that women were indeed more risk averse than men, consistent with the conventional wisdom.

This type of approach, in which only one covariate is *collected* and a “demographic effect” declared to be significant, is actually quite common in the experimental literature. Indeed, it is an appropriate approach if one is interested in total (or stratified) effect of a demographic characteristic rather than the marginal effect. Thus the estimated effect of sex, in this case, tells us the effect of sex itself *as well as all of the other characteristics of individuals that happen to be correlated with sex*. If women are smarter, older, and poorer than men, then these characteristics might be driving the result that there is a gender effect. Thus any individual that was smarter, older and poorer than the average would show the same effect, no matter what their sex. And in an important policy sense, there is indeed a gender effect: if one takes all of the characteristics of women together, as they come to us in the data, we observe a difference. For many policy purposes, this type of analysis is exactly the correct thing to undertake, as argued by Harrison, Lau and Williams [2002; p.1614]. The danger, of course, is that one intends to undertake an evaluation of the *marginal* effect of sex on risk attitudes, and incorrectly infers that from total effects such as these.

However, this sex effect appears to be a fragile sex effect when we allow for unobserved heterogeneity in behavior. Thus, even if one is only interested in the total effect, it is critical to account for heterogeneity in the population.

We extend the earlier random coefficient specification to allow sex to affect the mean of the population parameters (panel B.II of Table 1). In effect there are now separate means for men and women, and a common standard deviation.<sup>15</sup> Figure 4 illustrates: we are allowing the distributions shown in the bottom panel, compared to the distribution shown in the top panel. The results in Table 1 show that the sex effect is becoming statistically insignificant compared to the ML specification in panel A.II. The point estimate with random coefficients in Panel B.II implies a change in average risk aversion for the population of -0.186 when women make choices, but this

---

<sup>15</sup> It is a straightforward extension to allow the population standard deviation to also be a function of sex, in which case one could also estimate the covariance between the two population hyper-parameter estimates.

effect on the mean has a standard error of 0.128 and a  $p$ -value of 0.147. So the estimated size of the effect is less than half the estimated size when one undertakes the conventional analysis (panel A.II), and the  $p$ -value no longer meets conventional levels of statistical significance. More important than that, however, is the fact that the population standard deviation in risk attitudes is estimated to be 0.454, swamping the size of the effect from sex on the population mean. The gender effect is hard to discern given the level of heterogeneity in the population distribution of risk attitudes. Note that this is not saying that the estimates are “noisy” in the statistical sense of having large standard errors: the estimate of the standard deviation of the population can remain large even when samples are asymptotic.

We can allow for even greater heterogeneity from observables by allowing the mean to vary deterministically with additional demographic characteristics. Our discussion of the effect of this extension of the original random coefficient specification to just include sex illustrated the basic point, that these extensions will lead to shifts in the mean of the population distribution, but will not change the standard deviation of the population distribution; contrast the top panel of Figure 2 with the bottom panels. Figure 3 illustrates the possible effects for just two covariates. In our analysis we consider four covariates, apart from sex: the effect of dummy variables indicating if the subject reported being Black, Hispanic, having a Business major, or having a low GradePoint Average (below  $3\frac{1}{4}$ ).

The estimates in panel A.III show the effect of allowing for these covariates with a completely *non-random* coefficients specification, and then B.III shows the effect when we allow the intercept for the risk aversion coefficient to be random. The effect of sex remains statistically significant in the non-random coefficient specification in panel A.III, with women having a relative risk aversion that is 0.271 lower on average, with a standard error of only 0.158 and a 95% confidence interval between -0.581 and 0.038. On the other hand, the estimates in panel B.III of Table 1 show that sex has a less reliable effect on the average risk attitude when one allows for



random coefficients. The effect on the population mean is comparable,<sup>16</sup> but the estimated population standard deviation remains large in comparison.

The improvements in calculated log-likelihood as we expand the specification to allow for covariates and/or random coefficients is also shown in Table 1. Likelihood ratio tests confirm the intuition that these are statistically significant improvements as one compares similar ML and MSL specifications.<sup>17</sup> The same cannot be said for the comparisons of the alternative MSL specifications.

#### 4. The Relative Contribution of Observables and Unobservables

We can use these results to draw conclusions about the relative contribution of observable characteristics and unobservable heterogeneity. In one important sense *there is no comparison: the best specification includes both*. Thus panel B.III in Table 1 would be our “preferred specification” from the alternatives available, given these data and the inferential objective of characterizing individual heterogeneity with some insight into the determinants of that heterogeneity.

In terms of overall likelihood values, the “naked” random coefficient approach alone (panel B.I) outperforms the “dressed up” approach that relies solely on non-random covariates (panel A.III). On the other hand, the random coefficient approach by itself is silent on the determinants of heterogeneity.

For some inferential purposes that is of no significance, and in those instances the random coefficients approach is obviously attractive. For example, in the analysis of data from natural experiments, such as television game shows, one rarely has access to more than one or two individual characteristics (see Andersen, Harrison, Lau and Rutström [2008b] for a survey of the

---

<sup>16</sup> It is worth noting that these estimates can be numerically fragile for “small” numbers of Halton draws. If one uses 50 Halton draws, a common default, the coefficient for sex is estimated to be -0.124 with a  $p$ -value of 0.308. As one increases the number of Halton draws to 100, 200 and 300 this estimate ( $p$ -value) changes to -0.153 (0.252), -0.176 (0.163) and -0.193 (0.133), respectively. The estimates are numerically robust for Halton draws of 200 or higher, and all of the estimates we report in Table 1 use 500 Halton draws.

<sup>17</sup> Formally, the likelihood estimates in panel A of Table 1 cannot be used for likelihood ratio tests because they employ clustering, which violates the assumptions of the test. However, identical point estimates and likelihoods are obtained using MSL and constraining  $\sigma_{sd}$  to be zero, as one would expect given the formulae underlying the calculation of the robust standard errors in panel A of Table 1 (the standard errors are generally larger than those shown in Table 1). The likelihood ratio tests we report refer to these estimates, which may be validly employed.

literature on estimating risk attitudes from game shows). Similarly, in many experiments one might want to ascertain the effect of an experimental treatment, but there were no individual characteristics collected. Or, as noted earlier, the experimental study collected only one characteristic, such as sex or race, and seeks to make inferences about the marginal effect of that characteristic on behavior.

On the other hand, the demographic determinants of heterogeneity are often *the* focus of inference. In that case one can have a combination of approaches, as illustrated in panel B.III, and include observable characteristics in a random coefficients specification.

The conclusion we draw is perhaps common sense, that one should use observable covariates as well as flexible econometric specifications for unobserved heterogeneity whenever possible. Our approach makes that specification tractable for non-linear mixed logit models.

## **5. Modeling Extensions**

Our methodological goal is served by the simple, standard specifications we considered. But nothing in our approach, other than the precise quantitative conclusions, would be changed if one considered a range of extensions.

We could incorporate behavioral errors, to reflect the noise that experimental economists have come to expect in observed choice behavior. These extensions are standard, and well documented in the literature (e.g., Hey [1995][2005], Loomes and Sugden [1995][1998], Harrison and Rutström [2008; §2.3], and Wilcox [2008][2009]).

It is also possible to extend the theoretical specification to consider more exotic functional forms, and non-traditional specifications of choice under uncertainty such as Rank-Dependent Utility and Cumulative Prospect Theory. Nothing in our general approach mandates the use of the specific functional forms and models of choice we employ.

Each of these extensions raises the issue of allowing for multivariate random coefficient distributions. The approach, and software, we have developed allows these extensions, but our experience has been that one then undertake significant numerical checks of the reliability of

estimates. Of course, this is something that is capital-intensive and that can be automated using procedures such as those we employ in an appendix.

## 6. Additional Applications

Two additional applications illustrate the deeper significance of the non-linear mixed logit specification. The first is to stated choice responses widely used in marketing and environmental valuation, and the second is to the characterization of “subjective uncertainty.”

### *A. Stated Choice*

In marketing and environmental valuation the canonical choice tasks (e.g., Louviere, Hensher and Swait [2000]) often seem to require the respondent to make choices based on the evaluation of functionals that are best characterized as non-linear. For example, assume the attributes in question are lotteries: do you want brand X wine at price  $P_X$  that has 10% chance of being enjoyable and 90% chance of being horrid, or brand Y wine at prize  $P_Y (> P_X)$  that has a 50-50 chance of being enjoyable or horrid? Obviously indirect utility can be defined over prices, and income, but how do we capture the effect of the uncertainty? Perhaps by adding, with enough attributes of this general format, some *statistic* of the distribution of characteristics such as their standard deviation? But then we have implicitly assumed a Quadratic Utility function, which is restrictive. We cannot add the probabilities themselves, since there could be as many of these as there are attributes. This generic problem assumes policy significance when we think of applications of stated choice to serious environmental or health choices, where the uncertainty of final outcomes is endemic, and we need some way to untangle preferences over final outcomes from risk preferences.

Our non-linear mixed logit specification promises a resolution of this problem, allowing one to tease apart the role of risk preferences and innate preferences over attributes. The experimental or survey design would have to include some additional task that can be used to identify risk

preferences, and then make the maintained assumption that the risk preferences in that additional task apply equally to the stated choice task of inferential interest. For examples of joint estimation of this kind, exploiting the ability of experimental economists to add additional, identifying tasks to facilitate inference about a wide range of behavior, see Andersen, Harrison, Rutström and Lau [2008a] and Harrison and Rutström [2008; p.100-111].

### *B. Subjective Probability Distributions*

Subjective probabilities are widely viewed as being less certain than objective probabilities. In part this derives from the need to account for the observed preferences, positive or negative, that some subjects have for ambiguous settings, and more generally for settings in which subjective judgement plays a significant role. In part it derives from intuition that suggests that subjectively perceived chances of an event occurring might suffer from some lack of precision, and that this should matter for behavior.

Characterizing this presumed uncertainty is not easy, however. In some popular, traditional models of choice behavior, the uncertainty about the mean of a subjective probability distribution does not affect the observed choice behavior for an individual. In other words, if one person has a mean-preserving spread in their subjective beliefs, they would have no reason to change their observed choices when betting on the eventual outcome.

There are many ways to relax the traditional model and obtain potentially refutable predictions about how such changes in belief would affect behavior. Virtually all of these involve allowing for the subjective probability to be characterized as being uncertain, as if it is one draw from a subjective probability distribution (Gilboa, Postlewaite and Schmeidler [2008]). Hence one talks about subjective beliefs, rather than a single subjective probability. Anscombe and Aumann [1963], for example, proposed a characterization that presumed that the subjective probability of some event occurring was in fact a compound lottery of that event occurring with some probability and *then* rewarding the individual with some virtual lottery with uncertain payoffs. Hence the

subjective probability is actually a probability distribution, and not a scalar. One can then consider a range of non-traditional models that admit of refutable changes in observable behavior (e.g., Segal [1987], Gilboa and Schmeidler [1989], Klibanoff, Marinacci and Mukerji [2005] and Nau [2006]).

The non-linear mixed logit specification provides a direct method for estimating subjective beliefs, viewed as a subjective probability distribution. The key insight is to *characterize beliefs as a parameter to be estimated from observed choices in a well-defined experimental task, and to estimate that parameter as a random coefficient*. The experimental task consists of a series of standard lottery choices in which the subject is assumed to use conventional risk attitudes to select one lottery or the other, and then a series of betting choices in which the subject is presented with a range of bookies offering odds on some outcome that the subject has a belief over. The canonical event in the theoretical literature is the humble bingo cage, populated with a known number of balls that are orange and white, say. The subject knows that there are a certain number of balls in total, and that there are only orange and white balls. But the cage is only visible for about 10 seconds, and is rotating all the time to make it (practically) impossible for the subject to count the number of orange balls.

Our proposal is to view the subjective belief as a normal distribution in the population. There are two, equivalent ways that one could interpret these estimates. For the moment, to focus on essentials, assume that we have a representative decision maker in the population with homogenous beliefs about the event that are normally distributed. In fact we will allow for observable characteristics to weaken this assumption to some extent.

One interpretation is that the subjects pick one specific subjective *probability* value from that *distribution* when they place their bets: it is as if they draw a specific subjective probability from the population distribution “urn” and then use that probability draw to place their bets. In effect, one subject sees the turning bingo cage and thinks he sees 15% orange balls, another subject sees the same cage and thinks he sees 20%, and another subject sees the same cage and thinks he sees 25%; in each case they can be viewed as making draws from a single population distribution. Each subject then acts as if there is no uncertainty around those individual perceptions when making their bets.

This interpretation is exactly the one that the traditional econometric literature makes when estimating random coefficients, adapted for our task of estimating subjective beliefs.

Another interpretation is that our subjects each carry around with them a little, personal subjective probability distribution after they see the spinning bingo cage, and access that distribution when making their bets. This distribution can be assumed to be a replica of the population distribution, for simplicity sake, although that is a stronger assumption than is needed. Under this interpretation the estimated population distribution is in fact the distribution used by each subject.

Under either interpretation, the application of the non-linear mixed logit specification is immediate, and closes the empirical loop in the inference of subjective probability *distributions*.

## **7. Conclusions**

We develop a non-linear mixed logit specification that allows the analyst to directly estimate the “deep parameters” of interest in many behavioral analyses. We illustrated by considering the canonical case of estimating risk preferences from a sample of lottery choices. The data we consider are generated in controlled laboratory experiments, to ensure that we know all variables of interest, but the broader relevance of this issue is obvious.

**Table 1: Estimates**

Parameter	Coefficient	Point Estimate	Standard Error	p-value	95% Confidence	
A. Maximum Likelihood Estimates of Non-Linear Logit Model, Corrected for Clustering						
I. No Heterogeneity (LL= -2516.4)						
r	Constant	0.539	0.083	0.000	0.375	0.702
II. One Observed Covariate (LL= -2505.6)						
r	Constant	0.696	0.093	0.000	0.513	0.879
	Female	-0.321	0.126	0.011	-0.567	-0.074
III. More Covariates (LL= -2500.5)						
r	Constant	0.765	0.137	0.000	0.497	1.033
	Female	-0.271	0.158	0.086	-0.581	0.038
	Black	-0.193	0.220	0.381	-0.625	0.239
	Hispanic	-0.337	0.246	0.171	-0.818	0.145
	GPA <sub>low</sub>	-0.032	0.171	0.852	-0.367	0.303
	Business	-0.033	0.163	0.841	-0.353	0.288
B. Maximum Simulated Likelihood Estimates of Non-Linear Mixed Logit Model						
I. Unobserved Heterogeneity (LL= -2394.1)						
r <sub>mean</sub>	Constant	0.532	0.071	0.000	0.393	0.671
r <sub>sd</sub>	Constant	0.450	0.049	0.000	0.354	0.547
II. One Observed Covariate plus Unobserved Heterogeneity (LL= -2393.1)						
r <sub>mean</sub>	Constant	0.632	0.094	0.000	0.448	0.816
	Female	-0.186	0.128	0.147	-0.438	0.065
r <sub>sd</sub>	Constant	0.454	0.050	0.000	0.356	0.552
III. More Covariates plus Unobserved Heterogeneity (LL= -2390.7)						
r <sub>mean</sub>	Constant	0.682	0.122	0.000	0.442	0.922
	Female	-0.205	0.130	0.116	-0.461	0.051
	Black	-0.377	0.242	0.119	-0.853	0.097
	Hispanic	-0.321	0.220	0.144	-0.751	0.110
	GPA <sub>low</sub>	-0.053	0.128	0.676	-0.304	0.197
	Business	0.096	0.128	0.454	-0.156	0.349
r <sub>sd</sub>	Constant	0.423	0.048	0.000	0.329	0.517

Figure 1: Alternative Inferences About Population Parameter

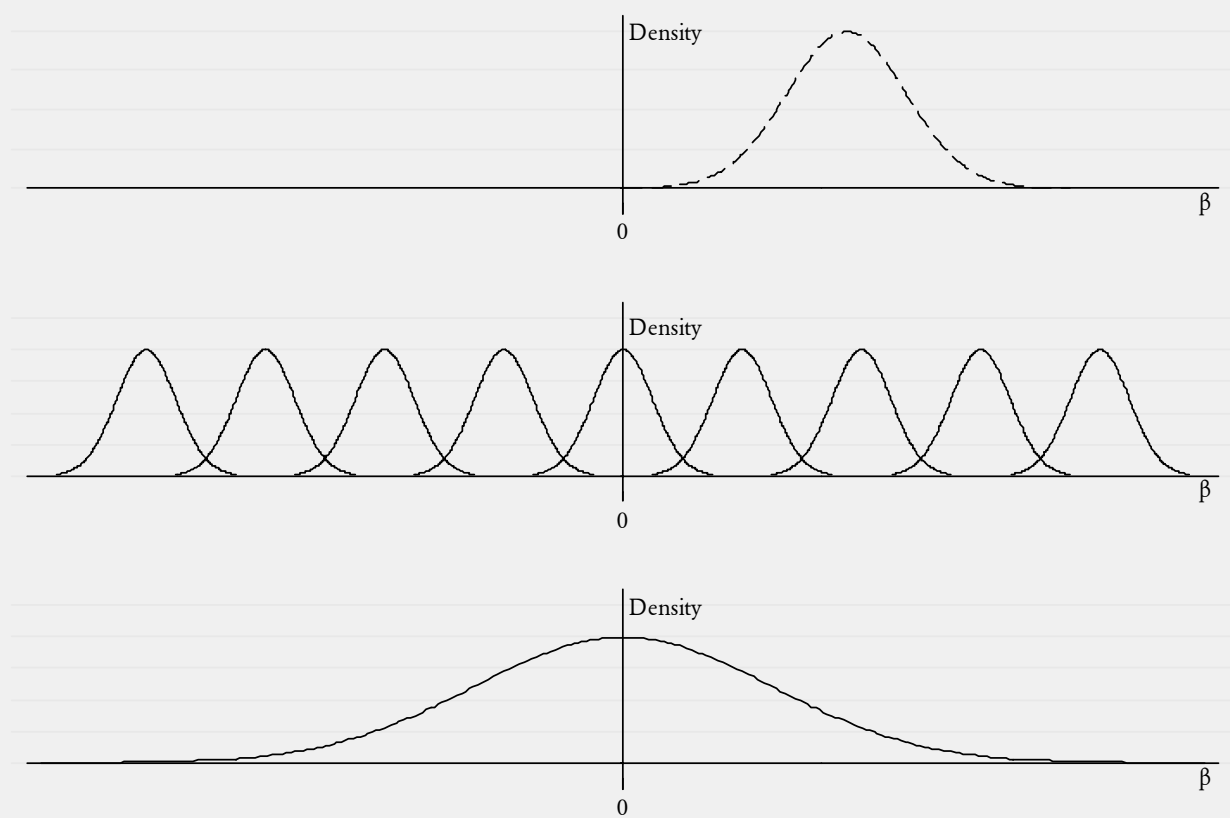




Figure 2: Effect on Different Random Coefficient Specifications

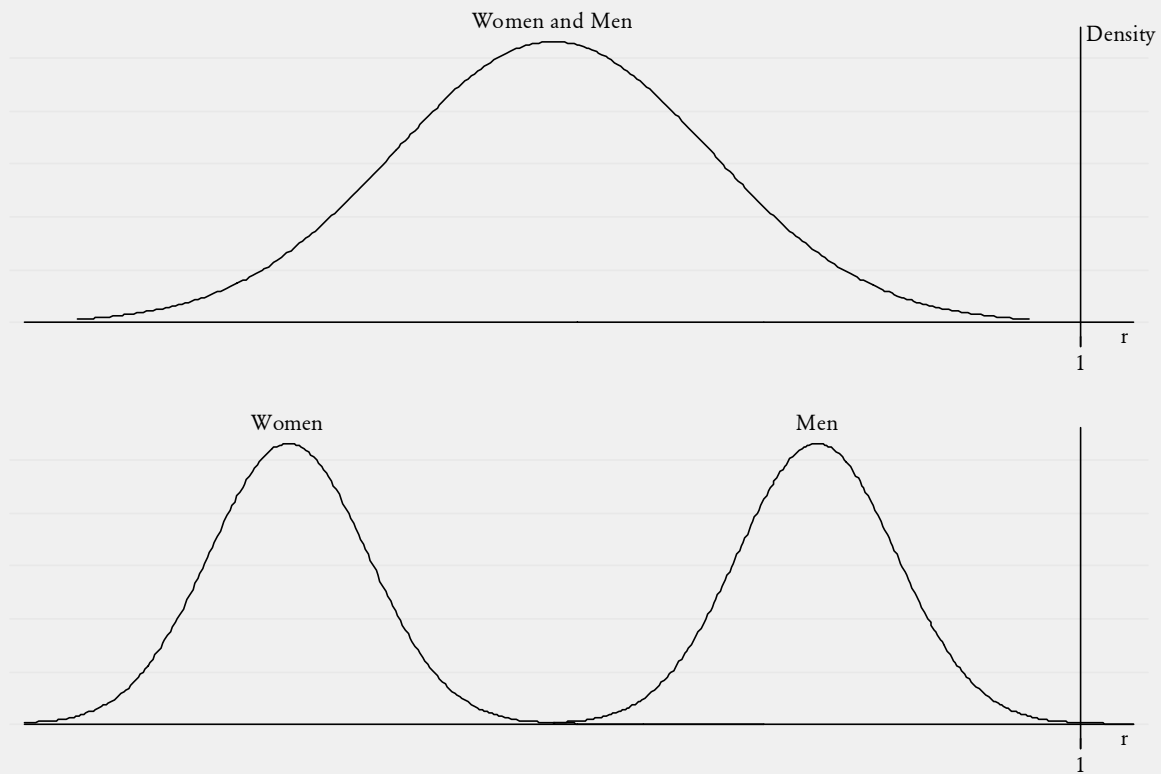
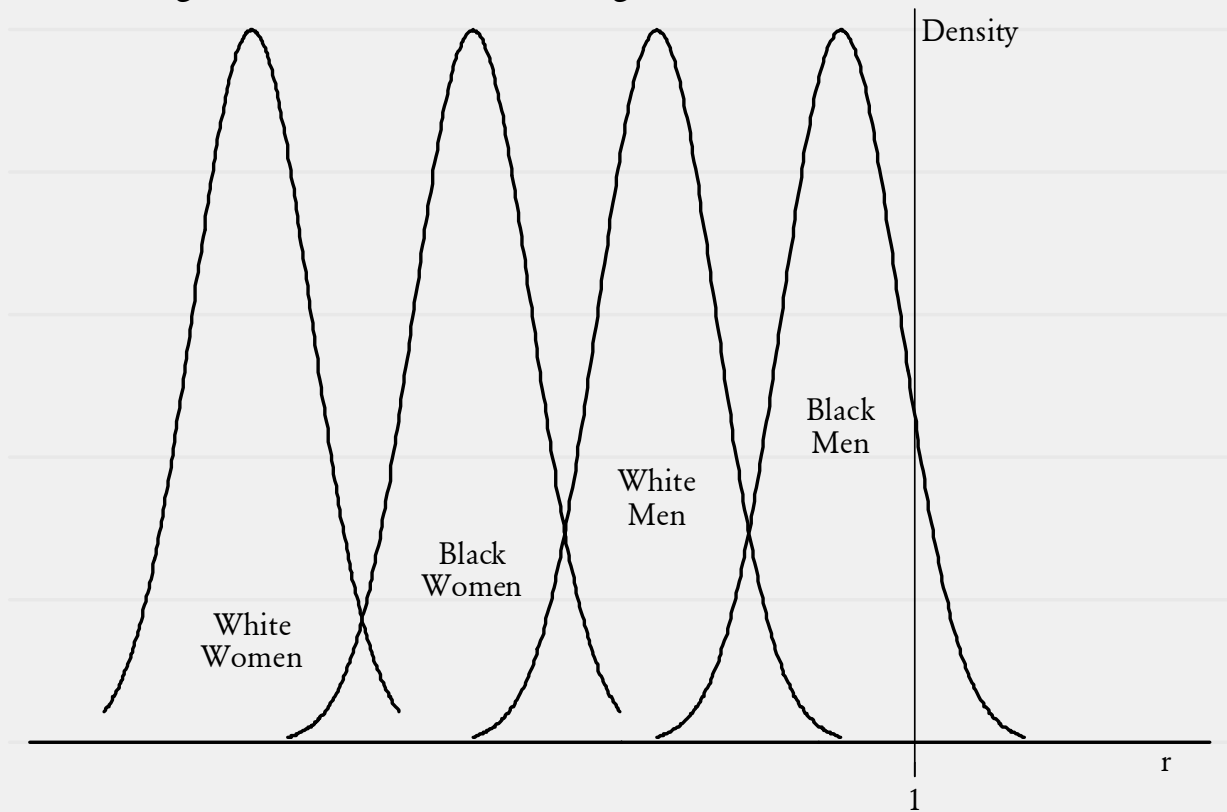


Figure 3: Effect of Allowing a Non-Random Covariate



## References

- Andersen, Steffen; Harrison, Glenn W.; Lau, Morten Igel, and Rutström, E. Elisabet, "Eliciting Risk and Time Preferences," *Econometrica*, 76(3), May 2008a, 583-618.
- Andersen, Steffen; Harrison, Glenn W., Lau, Morten I., and Rutström, E. Elisabet, "Risk Aversion in Game Shows," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emeraldm, Research in Experimental Economics, Volume 12, 2008b).
- Anscombe, F.J., and Aumann, R.J., "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 1963, 199-205.
- Botti, Fabrizio; Conte, Anna; DiCagno, Daniela, and D'Ippoliti, Carlo, "Risk Attitude in Real Decision Problems," *Advances in Economic Analysis & Policy*, 8(1), 2008, Article 6.
- Cabrales, Antonio, and García-Fontes, Walter, "Estimating Learning Models from Experimental Data," *Economics Working Paper 501*, Department of Economics and Business, Universitat Pompeu Fabra, September 2000.
- Conte, Anna; Hey, John D., and Moffatt, Peter G., "Mixture Models of Choice Under Risk," *Discussion Paper No. 2007/06*, Department of Economics and Related Studies, University of York, 2007.
- De Roos, Nicolas, and Sarafidis, Yianis, "Decision Making Under Risk in *Deal or No Deal*," *Working Paper*, School of Economics and Political Science, University of Sydney, April 2006.
- Drukker, David M., and Gates, Richard, "Generating Halton sequences using Mata," *Stata Journal*, 6(2), 214-228.
- Eckel, Catherine C., and Grossman, Philip J., "Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk," *Evolution and Human Behavior*, 23(4), 2002, 281-295.
- Gilboa, Itzhak; Postlewaite, Andrew P., and Schmeidler, David, "Probability and Uncertainty in Economic Modeling," *Journal of Economic Perspectives*, 22(3), Summer 2008, 173-188.
- Gilboa, Itzhak, and Schmeidler, David, "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18, 1989, 141-153.
- Harrison, Glenn W.; Lau, Morten I., and Rutström, E. Elisabet, "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109(2), 2007, 341-368.
- 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).
- Hensher, David A., and Greene, William H., "The Mixed Logit Model: The State of Practice," *Transportation*, 30, 2003, 133-176.
- 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, "Experimental Investigations of Errors in Decision Making Under Risk," *European Economic Review*, 39, 1995, 633-640.
- Hey, John D., "Why We Should Not be Silent About Noise," *Experimental Economics*, 8, 2005, 325-345.
- 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.
- Klibanoff, Peter; Marinacci, Massimo, and Mukerji, Sujoy, "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73(6), November 2005, 1849-1892.
- Kumbhakar, Subal C. and Lovell, C.A. Knox, *Stochastic Frontier Analysis* (New York: Cambridge University Press, 2000).
- Loomes, Graham, and Sugden, Robert, "Incorporating a Stochastic Element Into Decision Theories," *European Economic Review*, 39, 1995, 641-648.
- Loomes, Graham, and Sugden, Robert, "Testing Different Stochastic Specifications of Risky Choice," *Economica*, 65, 1998, 581-598.
- Louviere, Jordan J.; Hensher, David A., and Swait, Joffre D., *Stated Choice Methods: Analysis and Application* (New York: Cambridge University Press, 2000).
- McFadden, Daniel, and Train, Kenneth, "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 2000, 447-470.
- Nau, Robert F., "Uncertainty Aversion with Second-Order Utilities and Probabilities," *Management Science*, 52, 2006, 136-156.
- Revelt, David, and Train, Kenneth, "Mixed Logit With Repeated Choices: Households' Choices of Appliance Efficiency Levels," *Review of Economics & Statistics*, 80, 1998, 647-657.
- Segal, Uzi, "The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach," *International Economic Review*, 28(1), February 1987, 175-202.
- Train, Kenneth E., *Discrete Choice Methods with Simulation* (New York: Cambridge University Press, 2003).
- Wilcox, Nathaniel T., "Theories of Learning in Games and Heterogeneity Bias," *Econometrica*, 74(5), September 2006, 1271-1292.
- 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).
- Wilcox, Nathaniel T., "Stochastically More Risk Averse: A Contextual Theory of Stochastic Discrete Choice Under Risk," *Journal of Econometrics*, 142, 2009 forthcoming.

## Appendix A: 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. In the text we claim that the use of non-linear latent indices for utility, such as the ones we use, build 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 discussion 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  $\Delta$  from alternative  $j$  in task  $t$ , given by

$$\Delta_{njt} = \beta_n x_{njt} + \epsilon_{njt} \quad (A1)$$

where  $\beta_n$  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  $\epsilon_{njt}$  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  $z$  include the sex of the individual.

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

$$\Delta_{nA} = \eta_n w_{nA} + \gamma_n z_n + \epsilon_{nA} \quad (A2)$$

$$\Delta_{nB} = \eta_n w_{nB} + \gamma_n z_n + \epsilon_{nB} \quad (A3)$$

If we observe choice  $A$  then this choice reveals that

$$\Delta_{nA} > \Delta_{nB} \quad (A4)$$

Hence we have

$$\eta_n w_{nA} + \gamma_n z_n + \epsilon_{nA} > \eta_n w_{nB} + \gamma_n z_n + \epsilon_{nB} \quad (A5)$$

or

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

and therefore

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

So the vector  $z$  has dropped out completely.

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

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

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

$$\Delta_{nA} = H(\eta_n, \gamma_n, w_{nA}, z_n) + \varepsilon_{nA} \quad (A2')$$

$$\Delta_{nB} = H(\eta_n, \gamma_n, w_{nB}, z_n) + \varepsilon_{nB} \quad (A3')$$

Thus choice observation (A4) only implies

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

as a counterpart to (A5), and

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

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

### Additional References

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

## Appendix B: Software

We have developed a suite of programs in *Stata* to implement the non-linear mixed logit model. These programs extend existing programs and routines in *Stata*, in particular the **mixlogit** command developed and documented by Hole [2007]. In turn, all of these programs utilize lower-level routines within *Stata* for the efficient generation of random replicates using Halton or Hammersley sequences (Drukker and Gates [2006]). To facilitate the use of the non-linear mixed logit specification for the estimation of risk attitudes, we have written one special-purpose program called **mslcrra**, and one general-purpose program called **mixlognl**.

The special-purpose version only estimates constant relative risk aversion (CRRA) specifications assuming expected utility theory (EUT), but has the advantage of not requiring the user to write any code in the *Stata* Mata language: this is a language within *Stata* designed for efficient manipulation of matrices, but can initially be a bit cryptic to some users. We do provide options for the most commonly used variants, so the user can specify if the power CRRA function

$$U(m_k) = m_k^r \tag{B1}$$

is to be estimated, or the version used in many applied studies:

$$U(m_k) = m_k^{1-r}/(1-r) \tag{B2}$$

In addition, the user can specify with an option if a Fechner error term is to be included or not.

Various options for random draws are available, as in the general version.

The general version allows the user to specify any latent index that they like, and therefore allows direct extensions to the assumptions of CRRA or EUT theoretical assumptions. The cost is a modest investment in learning some basics of coding in Mata. To mitigate these costs we provide templates for most of the common specifications, and all of the variants estimated in the text. We discuss these templates below, and provide a gentle introduction to the necessary Mata syntax. It should be a relatively easy matter to then adapt these templates, if needed, for more exotic variants. Or one could use our special-purpose program as a template for other special-purpose models which are commonly used, such as rank-dependent utility (RDU) or cumulative prospect theory

(CPT) specifications. We prefer, for those extensions, to just write out the general-purpose template, but the objective is to provide programs that others can extend as appropriate for their needs.

The special purpose and general commands are written as programs in *Stata* ADO files, which are “read in” when called. For most users they simply need to include the relevant ADO files in the work directory that they are running their program in (the program that calls the command), or in a standard directory reserved for ADO files (and that *Stata* looks in if it cannot find the ADO file in the work directory). Specifically, the files `mslcr.ra.ado` and `mslcr.ra_d0.ado` execute the special-purpose command, and the files `mixlognl.ado`, `mixlognl_do.ado` and `mixlcov_nl.ado` execute the general command as well as a post-estimation “wrapper” program to display the estimated covariance matrix and/or standard deviations. Some lower-level routines are called as well, and are included in version 9.2 of *Stata* and later versions.

The examples discussed below are all contained in one *Stata* DO file called `msl_demo.do`. The commands to undertake the estimates reported in the text are in a *Stata* DO file called `msl.do`.

### *B1. The Special Purpose Commands*

The special purpose command is invoked with a line such as the following, to generate the “no covariates” specification, of the type reported<sup>18</sup> in Panel B of Table 1 of the text:

```
mslcr.ra choice, group(gid) id(id) rand(r_cons) prizes($prizes) probs($probs) search  
utype(1)
```

The variable `choice` denotes which lottery was chosen, and is typically binary in our applications; the `group` option, `gid` here, identifies the “choice situations” (explained below) faced by the subject; the `id` option, `id` here, identifies the individual making the choices; the `rand` option provides a label for the random variable, in this case `r_cons` instead of the default *Stata* label `_cons`; the `prizes` and `probs` options provide the names of the variables containing the data on the monetary prizes and

---

<sup>18</sup> The specific estimates reported in Table 1 may differ from those shown here, due to the use of significantly more Halton draws for the results in the text. Our examples in this appendix typically only use the default 50 Halton draws, which is fine for initial evaluations of models, but which can be unreliable. See Appendix C for more details on the type of numerical checks we advise.



probabilities of each lottery, and are given by the globals \$prizes and \$probs here (illustrated below); the search option tells the estimator to search for initial, feasible values; and the utype option specifies which utility function to use (1 refers to the power specification (B1), and 2 to the alternative specification (B2)). The globals are defined by statements such as

```
global prizes "prize0 prize1 prize2 prize3"
global probs  "prob0 prob1 prob2 prob3"
```

where the labels inside the quotes refer to variables defined as data. The “choice situations” in this setting refer to different choices for the individual. A binary lottery choice would therefore entail two rows of data, with the same gid value: the first row might refer to the left-hand lottery in the screen display, and the second row might refer to the right-hand lottery in the screen display. The choice variable takes on the value 1 for the row, and hence lottery, actually chosen, and the gid variable keeps track of which lotteries were being compared when that choice was made. This construction allows for the possibility, common in the area of stated choice, to allow for there to be more than two alternatives to be chosen from. Thus it is possible for this framework to estimate random-coefficient versions of multinomial logit models as well (see Long and Freese [2006; §7.2.4] or Hole [2007; p. 397ff] for illustrations of the manner in which this is done).

This example does not illustrate all of the options of the command. We come to those in a moment. The result of running the previous command will be a display such as this:

```
initial:      log likelihood = -2589.5979
alternative:  log likelihood = -2397.5403
rescale:      log likelihood = -2397.5403
rescale eq:   log likelihood = -2397.5403
Iteration 0:  log likelihood = -2397.5403
Iteration 1:  log likelihood = -2397.197
Iteration 2:  log likelihood = -2397.1814
Iteration 3:  log likelihood = -2397.1814
```

MSL CRRA model		Number of obs	=	7472
Log likelihood = -2397.1814		Wald chi2(1)	=	56.01
		Prob > chi2	=	0.0000

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Mean	r_cons	.5488889	.0733425	7.48	0.000	.4051403 .6926376
SD	r_cons	.5028208	.0582611	8.63	0.000	.3886311 .6170105

The estimates are displayed in standard *Stata* syntax, with the point estimates of the mean and standard deviation of the population distribution displayed along with their standard errors. Note that the typical label `_cons` has been replaced with the label `r_cons` so we know that it refers to the default value of the population parameter `r` in both the “Mean” and “SD” blocks of the display.

Most of the options are identical to those documented for the linear mixed logit command developed by Hole [2007]. The most important are the ones that control the specification of the random coefficients. The “nrep” field (e.g., `nrep(100)`) allows the choice of more or less Halton replicates than the default 50. The “burn” field (e.g., `burn(21)`) allows the choice of different numbers of burn sequences for the Halton draws, where the default is 15. The “search” field can be replaced with a “from” field, such as `from(start, copy)`, where the vector `start` is defined to be the starting values for the algorithm (e.g., one might define “`start = 0.549, 0.503`” after the initial estimation listed above has generated these values). And the “corr” field flags the estimator to allow non-zero correlations between random coefficients when there are two or more of them. This field would not be operative for the above specification, which assumes just one random coefficient; we illustrate its use momentarily.

In addition, there is a standard list of numerical algorithm options available. For our purposes the most important is to change the algorithm, in the “technique” field. Appendix C illustrates our use of this option to investigate the numerical reliability of MSL estimates.

To allow for two random coefficients one simply augments the “rand” field with the names of the variables to be used. We do not consider the multivariate random coefficient option in the text, but it may be of general interest so we illustrate it here. To allow for a male and a female random coefficient, one would use the dummy variable `Female` and incrementally change the command line to

```
mslcrra choice, group(gid) id(id) rand(Female r_cons) prizes($prizes) probs($probs)
      search utype(1)
```

This specification only changes the “rand” field, where we always leave the intercept label as the final label, per the usual *Stata* syntax of listing the intercept last. This implies that we assume zero

correlation between the two random coefficients, which is something we relax in a moment. This command generates the following output:

```
MSL CRRA model                                Number of obs   =       7472
Log likelihood = -2392.4358                    Wald chi2(2)    =       68.44
                                              Prob > chi2     =       0.0000
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Mean						
	Female	-.1430134	.1003021	-1.43	0.154	-.3396019 .0535751
	r_cons	.6207309	.0832154	7.46	0.000	.4576317 .7838302
SD						
	Female	.0879161	.0627636	1.40	0.161	-.0350982 .2109304
	r_cons	.473751	.0484194	9.78	0.000	.3788506 .5686513

So we read these estimates as saying that the mean of the `r_cons` random coefficient is 0.621, the mean of the Female random coefficient is -0.143, the standard deviation of the `r_cons` random coefficient is 0.473, and the standard deviation of the Female random coefficient is 0.088. And each of these point estimates, of course, has a standard error in the usual sense.

We extend the specification in an important manner by allowing there to be some correlation between the `r_cons` random coefficient and the Female random coefficient. The command line is virtually identical to the preceding, with the addition of “corr” at the end, and results in these estimates:

```
MSL CRRA model                                Number of obs   =       7472
Log likelihood = -2392.3516                    Wald chi2(2)    =       90.54
                                              Prob > chi2     =       0.0000
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	Female	-.2178263	.1156225	-1.88	0.060	-.4444423 .0087897
	r_cons	.663411	.0754491	8.79	0.000	.5155335 .8112885
	/111	.2153322	.0924822	2.33	0.020	.0340703 .396594
	/121	.3397455	.0502714	6.76	0.000	.2412153 .4382756
	/122	.2131554	.0446916	4.77	0.000	.1255616 .3007493

In fact, we made one other change to the command line. Instead of allowing *Stata* to search for starting values, we fed it the values from the preceding estimation in which the correlation was assumed zero. This was effected by the construction of vector `start`, and then instructing the command to use those values:

```
matrix tmp = e(b)
matrix start = tmp, 0
```

```
mslcrra choice, group(gid) id(id) rand(Female r_cons) prizes($prizes) probs($probs)
from(start, copy) corr utype(1)
```

The matrix e(b) contains the point estimates from the immediately preceding estimation exercise.

The display of estimates above is a little less transparent to read than the earlier examples. The first block shows the estimates of the “Mean” parameter, as before. So these point estimates change from -0.14 and 0.62 to be -0.22 and 0.66 when we allow for non-zero correlation. The second block now displays estimates of the lower-triangular elements of matrix L, where the estimated covariance matrix  $V = LL'$ . We recover the covariance matrix estimates, and the standard deviation estimates, using the post-estimation program `-mixlcov_nl-` we provided and these commands

```
mixlcov_nl
mixlcov_nl, sd
quietly mixlcov_nl, post
nlcom (correlation: _b[v21]/sqrt(_b[v11]*_b[v22]))
```

These commands result in the output

```
v11:  [111]_b[_cons]*[111]_b[_cons]
v21:  [121]_b[_cons]*[111]_b[_cons]
v22:  [121]_b[_cons]*[121]_b[_cons] + [122]_b[_cons]*[122]_b[_cons]
```

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
v11	.0463679	.0398288	1.16	0.244	-.0316951	.1244309
v21	.0731581	.0272502	2.68	0.007	.0197487	.1265676
v22	.1608622	.043217	3.72	0.000	.0761585	.2455659

for the covariance matrix, these estimates for the standard deviation

```
Female: sqrt([111]_b[_cons]*[111]_b[_cons])
r_cons: sqrt([121]_b[_cons]*[121]_b[_cons] + [122]_b[_cons]*[122]_b[_cons])
```

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Female	.2153322	.0924822	2.33	0.020	.0340703	.396594
r_cons	.4010763	.0538762	7.44	0.000	.2954808	.5066718

and these estimates for the correlation:

```
correlation: _b[v21]/sqrt(_b[v11]*_b[v22])
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
correlation	.8470843	.0533648	15.87	0.000	.7424913	.9516773

One pedagogically attractive feature of these displays is the explicit statement of the formulae used to calculate the items of interest, to make it reasonably transparent where they “come from” in terms of the initial block of estimates.

Taking stock of these final estimates, then, we calculate the mean of the `r_cons` random coefficient is 0.663, the mean of the Female random coefficient is -0.218, the standard deviation of the `r_cons` random coefficient is 0.401, the standard deviation of the Female random coefficient is 0.215, and the correlation between the `r_cons` and Female random coefficients is 0.847. Each of these point estimates, to repeat, has a corresponding standard error.

To include non-random coefficients one augments the initial part of the command in the usual syntax for *Stata* with “right hand side variable.” Thus we would issue the command

```
mslccrra choice Female Black Hispanic Business GPAlow, group(gid) id(id) rand(r_cons)
prizes($prizes) probs($probs) utype(1)
```

if we wanted to have non-random coefficients reflecting the data variables Female, Black, Hispanic, Business and GPAlow, as in Panel F of Table 1. These non-random coefficients apply to the `r_cons` random coefficient, as explained in the text.

The special-purpose command also has an option to include a “Fechner” error term, by simply adding the word **fechner** at the end of the command line. We define this specification below formally, and show how it affects the evaluation of the model. The Fechner error term in the special-purpose command is non-random, but in the general-purpose command it can also be a random coefficient.

## *B2. The General Commands*

The general command works in a similar manner, except that the non-linear function has to be specified explicitly in a Mata program. The command for estimating the random coefficient model with no covariates, and no Fechner error term, is

```
mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3, group(gid) id(id)
kfix(0) krnd(1) search
```

There are some small syntax changes from the special command. The data on choice responses is now a binary variable, to flag with a 1 the alternative selected by the subject (again, data are stacked in rows, one row per choice alternative per subject). The data on lottery prizes and probabilities are entered in the initial section of the command, rather than as an option. And we have two new fields, “kfix” and “krnd,” to identify how many variables are to be treated as non-random (or fixed) and how many are to be treated as random. In this instance we only have one random coefficient and no fixed (non-random) coefficients.

The major difference with the general command is the need to write a Mata function to define the latent process in the  $H(\cdot)$  function in expression (2) in the text. In our case this function is as follows:

```

mata: mata clear
mata:
function utilfunc(XMAT, BETA)
{
    // BETA = 1 x nrep matrix of coefficients
    // XMAT[1,1..4] = 1 x 4 matrix of probabilities for lottery 1
    // XMAT[2,1..4] = 1 x 4 matrix of probabilities for lottery 2
    // XMAT[1,5..8] = XMAT[2,5..8] = 1 x 4 matrix of prizes
    // Note that :* denotes element-by-element multiplication
    // Note that :^ denotes element-by-element exponentiation

    R = BETA[1,.]

    V1 = XMAT[1,1..4] :* (XMAT[1,5]:^R',XMAT[1,6]:^R',XMAT[1,7]:^R',XMAT[1,8]:^R')
    V2 = XMAT[2,1..4] :* (XMAT[2,5]:^R',XMAT[2,6]:^R',XMAT[2,7]:^R',XMAT[2,8]:^R')

    F1 = colsum(V1')
    F2 = colsum(V2')
    PEU = F1 \ F2

    EV = exp(PEU) ./ colsum(exp(PEU))

    return(EV)
}
end

```

Complete details on the Mata syntax used here can be obtained from the *Mata Matrix Programming Manuals* in StataCorp [2007], but the basic logic should be apparent. The purpose of this function is to take a row vector called BETA, and data on lottery prizes and probabilities in a matrix X, to generate a column vector EV that evaluates the function  $H(\cdot)$ . Note that “EV” is meant as a generic shorthand for “evaluation” rather than “expected value.” Each row of the matrix X refers to one of

the alternatives, so there are only 2 rows since we consider binary choice. The first 4 columns of the matrix X here refer to the lottery probabilities, and the last 4 columns refer to the lottery prizes, where the prizes and probabilities are ordered to match up with each other. The expected utility of each lottery is evaluated in V1 and V2, using the power utility function and assuming an EUT specification. The only novelty here is perhaps the syntax that Mata uses for element-wise multiplication and exponentiation, but that is documented. The remainder of the function implements the logic of our basic latent model.

The estimates obtained with this specification are as follows:

Mixed logit model				Number of obs	=	7472
				Wald chi2(0)	=	.
Log likelihood = -2397.1814				Prob > chi2	=	.
-----						
	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-----						
mean1						
	_cons	.5488889	.0733425	7.48	0.000	.4051403 .6926376
-----						
sdl						
	_cons	.5028208	.0582611	8.63	0.000	.3886311 .6170105
-----						

This replicates exactly the results of the special-purpose code. The display is a little more generic, as one might expect. In this case it is apparent what the coefficients refer to, but when there are more than one covariate this will require some (modest) attention, as illustrated below.

We can extend the Mata function to include a Fechner error term. 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 a review). 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 it 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  $\mu$  gets larger and larger, the predicted choice converges to indifference.

In our earlier notation, this amounts to a change in the specification that is captured in the non-linear  $H(\cdot)$  function in expression (2). Specifically, if we define

$$\Gamma_{njt} = EU_R(\beta_n, x_{njt}) - EU_L(\beta_n, x_{njt}) \quad (B3)$$

as the difference in EU of the right (R) and left (L) lotteries, then our previous specification

$$H(\beta_n, x_{njt}) = \exp(\Gamma(\beta_n, x_{njt})) / (1 + \exp(\Gamma(\beta_n, x_{njt}))) \quad (B4)$$

becomes

$$H(\beta_n, x_{njt}) = \exp(\Gamma(\beta_n, x_{njt})/\mu) / (1 + \exp(\Gamma(\beta_n, x_{njt})/\mu)) \quad (B5)$$

with the addition of the Fechner error parameter  $\mu$ , which is now a part of  $\beta_n$  along with the deep parameter  $r$ . We can treat the Fechner parameter as a random or non-random covariate, of course.

Our Mata implementation is as follows:

```

mata: mata clear
mata:
function utilfunc(XMAT, BETA)
{
    R = BETA[1,.]
    MU = exp(BETA[2,.])

    V1 = XMAT[1,1..4] :* (XMAT[1,5]:^R',XMAT[1,6]:^R',XMAT[1,7]:^R',XMAT[1,8]:^R')
    V2 = XMAT[2,1..4] :* (XMAT[2,5]:^R',XMAT[2,6]:^R',XMAT[2,7]:^R',XMAT[2,8]:^R')

    // Fechner error index, the extra beta
    F1 = colsum(V1') :/ MU
    F2 = colsum(V2') :/ MU
    PEU = F1 \ F2
    EV = exp(PEU) :/ colsum(exp(PEU))

    return(EV)
}

```

We augment the BETA vector by a new row, to reflect the fact that we now have two parameters: hence BETA is now a matrix with two rows. The MU parameter is treated as a log-normal term, to ensure that the internal value within this function is non-negative, while allowing the parameter driven by the outer algorithm to search between  $\pm\infty$  for good values. The EU terms, V1 and V2, are



calculated as before, and then simply divided by the value of MU. This implements a mathematically equivalent version of expression (B5). It is easy to show that (B5) is equivalent to  $\exp(EU_R(\beta_n, x_{njt})/\mu) / (\exp(EU_L(\beta_n, x_{njt})/\mu) + \exp(EU_R(\beta_n, x_{njt})/\mu))$ , which is the version we implement: multiply (B5) by  $\exp(EU_L(\beta_n, x_{njt})/\mu) / \exp(EU_L(\beta_n, x_{njt}))$  and simplify.

We can then loop through a range of specifications with this Mata function. Here we initiate the search at random each time, but in general one would use solution values from one stage to initiate the estimation at the next stage:

```
. constraint 1 [sd1]_cons = 0
. constraint 2 [sd2]_cons = 0

. * assume that both r and mu are non-random
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 , group(gid) id(id)
  kfix(0) krnd(2) search nrep(50) constraint(1 2)
```

```
Mixed logit model                      Number of obs   =       7472
Log likelihood = -2415.8271              Wald chi2(0)    =       .
                                          Prob > chi2     =       .
```

```
( 1) [sd1]_cons = 0
( 2) [sd2]_cons = 0
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
mean1							
	_cons	.467769	.0144263	32.42	0.000	.439494	.4960441
mean2							
	_cons	-1.509273	.0594646	-25.38	0.000	-1.625821	-1.392725
sd1							
	_cons	(dropped)					
sd2							
	_cons	(dropped)					

```
. * assume that r is random and mu is non-random
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 , group(gid) id(id)
  kfix(0) krnd(2) search nrep(50) constraint(2)
```

```
Mixed logit model                      Number of obs   =       7472
Log likelihood = -2196.403              Wald chi2(0)    =       .
                                          Prob > chi2     =       .
```

```
( 1) [sd2]_cons = 0
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
mean1							
	_cons	.2708522	.0828113	3.27	0.001	.108545	.4331593
mean2							
	_cons	-1.95416	.0600839	-32.52	0.000	-2.071922	-1.836397
sd1							
	_cons	.5613064	.0870761	6.45	0.000	.3906405	.7319723
sd2							

```

      _cons | (dropped)
-----+-----
. * assume that r and mu are random
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 , group(gid) id(id)
      kfix(0) krnd(2) search nrep(50)

Mixed logit model                                Number of obs   =       7472
Log likelihood = -2129.561                      Wald chi2(0)      =       .
                                              Prob > chi2       =       .

-----+-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
mean1   _cons |    .4601292   .027792    16.56   0.000    .4056578   .5146006
-----+-----
mean2   _cons |   -2.033982   .0911654   -22.31   0.000   -2.212663  -1.855301
-----+-----
sd1     _cons |    .4999122   .0401051    12.47   0.000    .4213078   .5785167
-----+-----
sd2     _cons |    .9479235   .1130785     8.38   0.000    .7262936   1.169553
-----+-----

. * assume that r and mu are random, and allow a correlation
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 , group(gid) id(id)
      kfix(0) krnd(2) search nrep(50) corr

Mixed logit model                                Number of obs   =       7472
Log likelihood = -2107.6966                      Wald chi2(0)      =       .
                                              Prob > chi2       =       .

-----+-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
mean1   _cons |    .4713754   .0362431    13.01   0.000    .4003403   .5424105
-----+-----
mean2   _cons |   -1.827599   .0904332   -20.21   0.000   -2.004845  -1.650353
-----+-----
      /111 |   -.3952561   .0364335   -10.85   0.000   -.4666645  -.3238478
      /121 |   -.666333   .1169786    -5.70   0.000   -.8956068  -.4370591
      /122 |   .4972033   .0942341     5.28   0.000    .3125077   .6818988

Mixed logit model                                Number of obs   =       7472
Log likelihood = -2107.6966                      Wald chi2(0)      =       .
                                              Prob > chi2       =       .

-----+-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
mean1   _cons |    .4713754   .0362431    13.01   0.000    .4003403   .5424105
-----+-----
mean2   _cons |   -1.827599   .0904332   -20.21   0.000   -2.004845  -1.650353
-----+-----
      /111 |   -.3952561   .0364335   -10.85   0.000   -.4666645  -.3238478
      /121 |   -.666333   .1169786    -5.70   0.000   -.8956068  -.4370591
      /122 |   .4972033   .0942341     5.28   0.000    .3125077   .6818988
-----+-----

. mixlcov_nl
      v11:  [111]_b[_cons]*[111]_b[_cons]
      v21:  [121]_b[_cons]*[111]_b[_cons]

```

```

v22:  [121]_b[_cons]*[121]_b[_cons] + [122]_b[_cons]*[122]_b[_cons]
-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    v11 |    .1562274   .0288011     5.42   0.000    .0997782    .2126766
    v21 |    .2633722   .0643835     4.09   0.000    .1371828    .3895616
    v22 |    .6912107   .1539468     4.49   0.000    .3894805    .9929409
-----

. mixlcov_nl, sd

      sd1:  sqrt([111]_b[_cons]*[111]_b[_cons])
      sd2:  sqrt([121]_b[_cons]*[121]_b[_cons] + [122]_b[_cons]*[122]_b[_cons])
-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
    sd1 |    .3952561   .0364335    10.85   0.000    .3238478    .4666645
    sd2 |    .8313908   .0925839     8.98   0.000    .6499297    1.012852
-----

```

The estimates are consistent with those obtained with the special-purpose program. Note that one has to keep track of which variable is which in terms of the display. The Fechner error term is always the last term, so in the above displays it is “mean2,” and the intercept is given by “mean1.” The estimates for mean1 and sd1 go together, as do the estimates for mean2 and sd2, etc.

We next illustrate how to augment the Mata function to allow for a “regular” covariate such as the dummy variable Female that is the substantive focus of our analysis.

```

mata: mata clear
mata:
function utilfunc(XMAT,BETA)
{
    R = XMAT[1,9]*BETA[1,.] :+ BETA[2,.]
    MU = exp(BETA[3,.])

    V1 = XMAT[1,1..4] :* (XMAT[1,5]:^R',XMAT[1,6]:^R',XMAT[1,7]:^R',XMAT[1,8]:^R')
    V2 = XMAT[2,1..4] :* (XMAT[2,5]:^R',XMAT[2,6]:^R',XMAT[2,7]:^R',XMAT[2,8]:^R')

    // Fechner error index, the extra beta
    F1 = colsum(V1') :/ MU
    F2 = colsum(V2') :/ MU
    PEU = F1 \ F2
    EV = exp(PEU) :/ colsum(exp(PEU))

    return(EV)
}
end

```

Here we now have 3 rows to the parameter vector BETA. The last row, as always, is the Fechner parameter  $\mu$ . The first row is the new covariate Female, and the second row is the old covariate for the intercept. After R and MU are defined, taking this into account, everything is the same as in the previous Mata function. Note that R is defined now as the linear function of the dummy variable

Female, which is stored internally in XMAT[1,9] and a constant. The coefficient for Female, BETA[1,] is therefore multiplied by XMAT[1,9] and a constant, BETA[2,] added element-wise. Thus one could easily augment this specification with additional random covariates.

The syntax for invoking the command now adds the variable Female, and modifies the treatment of covariates as random or fixed using one of two method:

```
* assume Fechner is non-random using constraint option, and Female is non-random using
  kfix and krnd options

constraint 3 [sd3]_cons = 0

mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
  id(id) kfix(1) krnd(2) search nrep(50) constraint(3)

* assume Fechner is non-random using constraint option, and Female is random using
  kfix and krnd options

mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
  id(id) kfix(0) krnd(3) search nrep(50) constraint(3)

* all three are random
mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
  id(id) kfix(0) krnd(3) search nrep(50)

* everything is random, with correlation allowed
mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
  id(id) kfix(0) krnd(3) search nrep(50) corr

mixlcov_nl
mixlcov_nl, sd
```

Note that one can use the constraint option to impose a zero standard deviation on some random coefficient, thereby constraining it to be fixed, or one can explicitly declare it to be fixed using combinations of the kfix and krnd options.

Here are estimation results from these specifications:

```
. * assume Fechner is non-random using constraint option, and Female is non-random
  using kfix and krnd options

. constraint 3 [sd3]_cons = 0

. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
  id(id) kfix(1) krnd(2) search nrep(50) difficult constraint(3)
```

Mixed logit model

Log likelihood = -2194.2318

Number of obs = 7472  
Wald chi2(0) = .  
Prob > chi2 = .

```
( 1) [sd3]_cons = 0
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
mean1						
	_cons	-.186698	.0812328	-2.30	0.022	-.3459113 -.0274846
mean2						
	_cons	.4099984	.0706687	5.80	0.000	.2714903 .5485064

mean3						
_cons	-1.945708	.0577978	-33.66	0.000	-2.058989	-1.832426
sd2						
_cons	.5256255	.0667201	7.88	0.000	.3948565	.6563945
sd3						
_cons	(dropped)					

```
. * assume Fechner is non-random using constraint option, and Female is random using
    kfix and krnd options
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
    id(id) kfix(0) krnd(3) search nrep(50) difficult constraint(3)
```

```
Mixed logit model                      Number of obs   =       7472
                                         Wald chi2(0)     =           .
Log likelihood = -2190.8985              Prob > chi2      =           .
```

```
( 1) [sd3]_cons = 0
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
mean1						
_cons		-.0539577	.0630112	-0.86	0.392	-.1774573 .069542
mean2						
_cons		.4671867	.0608082	7.68	0.000	.3480048 .5863687
mean3						
_cons		-1.908129	.0504182	-37.85	0.000	-2.006947 -1.809311
sd1						
_cons		.3423815	.0484568	7.07	0.000	.2474079 .4373552
sd2						
_cons		.3492558	.0407876	8.56	0.000	.2693135 .4291981
sd3						
_cons		(dropped)				

```
. * all three are random
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
    id(id) kfix(0) krnd(3) search nrep(50) difficult
```

```
Mixed logit model                      Number of obs   =       7472
                                         Wald chi2(0)     =           .
Log likelihood = -2119.5198              Prob > chi2      =           .
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
mean1						
_cons		-.3184332	.0490969	-6.49	0.000	-.4146612 -.2222051
mean2						
_cons		.5428351	.0339079	16.01	0.000	.4763769 .6092933
mean3						
_cons		-1.82915	.1135632	-16.11	0.000	-2.05173 -1.60657
sd1						
_cons		.0967169	.0358776	2.70	0.007	.0263981 .1670357
sd2						
_cons		.3843775	.029437	13.06	0.000	.326682 .442073
sd3						

```

      _cons | .8443798 .1062071 7.95 0.000 .6362176 1.052542
-----+-----
. * everything is random, with correlation allowed
. mixlognl y prob0 prob1 prob2 prob3 prize0 prize1 prize2 prize3 Female, group(gid)
      id(id) kfix(0) krnd(3) search nrep(50) difficult corr

Mixed logit model                                Number of obs   =       7472
                                                Wald chi2(0)       =           .
Log likelihood = -2103.6909                    Prob > chi2        =           .

-----+-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
mean1   _cons |   -.0503877   .057993    -0.87   0.385    - .1640519   .0632764
-----+-----
mean2   _cons |    .5146076   .0556593     9.25   0.000     .4055174   .6236978
-----+-----
mean3   _cons |   -1.852156   .1134007    -16.33   0.000    -2.074417  -1.629894
-----+-----
      /111 |   -.0928146   .0583093     -1.59   0.111    - .2070988   .0214695
      /121 |   -.2870913   .0481878     -5.96   0.000    - .3815376  - .1926449
      /131 |   -.2394348   .1405451     -1.70   0.088    - .514898    .0360285
      /122 |   -.2382972   .0249106     -9.57   0.000    - .2871212  - .1894733
      /132 |   -.7919702   .0880972     -8.99   0.000    - .9646375  - .6193029
      /133 |    .0248046   .1490321     0.17   0.868    - .2672929   .3169021
-----+-----

. mixlcov_nl

      v11:  [111]_b[_cons]*[111]_b[_cons]
      v21:  [121]_b[_cons]*[111]_b[_cons]
      v31:  [131]_b[_cons]*[111]_b[_cons]
      v22:  [121]_b[_cons]*[121]_b[_cons] + [122]_b[_cons]*[122]_b[_cons]
      v32:  [131]_b[_cons]*[121]_b[_cons] + [132]_b[_cons]*[122]_b[_cons]
      v33:  [131]_b[_cons]*[131]_b[_cons] + [132]_b[_cons]*[132]_b[_cons] +
[133]_b[_cons]*[133]_b[_cons]

-----+-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      v11 |    .0086146   .0108239     0.80   0.426    - .0125999   .029829
      v21 |    .0266463   .015693     1.70   0.090    - .0041114   .0574039
      v31 |    .022223    .0195416     1.14   0.255    - .0160777   .0605238
      v22 |    .139207    .0329341     4.23   0.000     .0746573   .2037567
      v32 |    .2574639   .0712204     3.62   0.000     .1178746   .3970533
      v33 |    .6851611   .1772148     3.87   0.000     .3378265   1.032496
-----+-----

. mixlcov_nl, sd

      sd0:  sqrt([111]_b[_cons]*[111]_b[_cons])
      sd1:  sqrt([121]_b[_cons]*[121]_b[_cons] + [122]_b[_cons]*[122]_b[_cons])
      sd2:  sqrt([131]_b[_cons]*[131]_b[_cons] + [132]_b[_cons]*[132]_b[_cons] +
[133]_b[_cons]*[133]_b[_cons])

-----+-----
      y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      sd0 |    .0928146   .0583093     1.59   0.111    - .0214695   .2070988
      sd1 |    .3731045   .0441353     8.45   0.000     .286601    .4596081
      sd2 |    .8277446   .1070468     7.73   0.000     .6179368   1.037552
-----+-----

```

Happily, the log-likelihood improves with every step towards a more flexible specification.

Finally, it might be useful to illustrate the changes in syntax needed to define different functional forms than the simple power-CRRA utility function, and EUT model, we have been using throughout. One popular option, following Holt and Laury [2002] and Saha [1993], is the Expo-Power utility function which allows increasing or decreasing relative risk aversion (RRA). Instead of

$$U(m_k) = m_k^r \quad (B6)$$

we would have

$$U(m_k) = [1 - \exp(-\alpha m_k^{1-r})] / \alpha, \quad (B7)$$

where  $\alpha$  and  $r$  are parameters to be estimated. RRA is then  $r + \alpha(1-r)m_k^{1-r}$ , so RRA varies with income  $m_k$  if  $\alpha \neq 0$ . This function nests CRRA (as  $\alpha \rightarrow 0$ ) and CARA (as  $r \rightarrow 0$ ). The Mata code is modified as follows to implement this specification, still within an EUT framework:

```
R = BETA[1,.]
ALPHA = BETA[2,.]
MU = exp(BETA[3,.])

V1 = (1:-exp(-ALPHA'*(XMAT[1,5]:^R'))):/ALPHA'
V1 = V1,(1:-exp(-ALPHA'*(XMAT[1,6]:^R'))):/ALPHA'
V1 = V1,(1:-exp(-ALPHA'*(XMAT[1,7]:^R'))):/ALPHA'
V1 = V1,(1:-exp(-ALPHA'*(XMAT[1,8]:^R'))):/ALPHA'
V1 = colsum((XMAT[1,1..4] :* V1)') :/ MU

V2 = (1:-exp(-ALPHA'*(XMAT[2,5]:^R'))):/ALPHA'
V2 = V2,(1:-exp(-ALPHA'*(XMAT[2,6]:^R'))):/ALPHA'
V2 = V2,(1:-exp(-ALPHA'*(XMAT[2,7]:^R'))):/ALPHA'
V2 = V2,(1:-exp(-ALPHA'*(XMAT[2,8]:^R'))):/ALPHA'
V2 = colsum((XMAT[2,1..4] :* V2)') :/ MU
```

The remaining code remains the same, once V1 and V2 are defined.

The Rank-Dependent specification of probability weights, due originally to Quiggin [1982], can be implemented as well, and illustrates how a popular alternative to EUT can be estimated. Most of the new coding is just to keep track of the rank-ordering of prizes, which is a crucial feature of this specification. The economics of this model are well-known, and our specification follows the exposition and (familiar) parametric forms adopted by Harrison and Rutström [2008; p.87]:

```
R = BETA[1,.]
GAMMA = BETA[2,.]
MU = exp(BETA[3,.])

// Now do the probability weights initially on the first lottery//
// Generate the cumulative probabilities //

P1 = 0,0,0,0
if (XMAT[1,4]!=0) P1[1,4] = XMAT[1,4]
if (XMAT[1,3]!=0) P1[1,3] = XMAT[1,3] + XMAT[1,4]
if (XMAT[1,2]!=0) P1[1,2] = XMAT[1,2] + XMAT[1,3] + XMAT[1,4]
```

```

if (XMAT[1,1]!=0) P1[1,1] = XMAT[1,1] + XMAT[1,2] + XMAT[1,3] + XMAT[1,4]

// Generate the weighted cumulative probabilities by each element in the P1 vector//
D1 = (P1[1,1]:^GAMMA'):/((P1[1,1]:^GAMMA'+(1:-P1[1,1]):^GAMMA'):^(1:/GAMMA'))
D1 = D1,(P1[1,2]:^GAMMA'):/((P1[1,2]:^GAMMA'+(1:-P1[1,2]):^GAMMA'):^(1:/GAMMA'))
D1 = D1,(P1[1,3]:^GAMMA'):/((P1[1,3]:^GAMMA'+(1:-P1[1,3]):^GAMMA'):^(1:/GAMMA'))
D1 = D1,(P1[1,4]:^GAMMA'):/((P1[1,4]:^GAMMA'+(1:-P1[1,4]):^GAMMA'):^(1:/GAMMA'))
// Note that D1 is a nrep x 4 matrix

// Generate the decision weights
// Expressions like (D1[.,3]:>0) evaluates to 1 when true and zero otherwise
W1 = J(cols(BETA),4,0)
W1[.,4] = D1[.,4]
W1[.,3] = D1[.,3] :- D1[.,4]*(D1[.,3]:>0)
W1[.,2] = D1[.,2] :- D1[.,3]*(D1[.,2]:>0) :- D1[.,4]*(D1[.,3]==0)*(D1[.,2]:>0)
W1[.,1] = D1[.,1] :- D1[.,2]*(D1[.,1]:>0) :- D1[.,3]*(D1[.,2]==0)*(D1[.,1]:>0) :-
    D1[.,4]*(D1[.,3]==0)*(D1[.,2]==0)*(D1[.,1]:>0)

// Now do the probability weights initially on the second lottery//
// Generate the cumulative probabilities //

P2 = 0,0,0,0
if (XMAT[2,4]!=0) P2[1,4] = XMAT[2,4]
if (XMAT[2,3]!=0) P2[1,3] = XMAT[2,3] + XMAT[2,4]
if (XMAT[2,2]!=0) P2[1,2] = XMAT[2,2] + XMAT[2,3] + XMAT[2,4]
if (XMAT[2,1]!=0) P2[1,1] = XMAT[2,1] + XMAT[2,2] + XMAT[2,3] + XMAT[2,4]

// Generate the weighted Cumulative probabilities by each element in the P2 vektor//
D2 = (P2[1,1]:^GAMMA'):/((P2[1,1]:^GAMMA'+(1:-P2[1,1]):^GAMMA'):^(1:/GAMMA'))
D2 = D2,(P2[1,2]:^GAMMA'):/((P2[1,2]:^GAMMA'+(1:-P2[1,2]):^GAMMA'):^(1:/GAMMA'))
D2 = D2,(P2[1,3]:^GAMMA'):/((P2[1,3]:^GAMMA'+(1:-P2[1,3]):^GAMMA'):^(1:/GAMMA'))
D2 = D2,(P2[1,4]:^GAMMA'):/((P2[1,4]:^GAMMA'+(1:-P2[1,4]):^GAMMA'):^(1:/GAMMA'))
// Note that D2 is a nrep x 4 matrix

// Generate the decision weights
// Expressions like (D2[.,3]:>0) evaluate to 1 when true and zero otherwise
W2 = J(cols(BETA),4,0)
W2[.,4] = D2[.,4]
W2[.,3] = D2[.,3] :- D2[.,4]*(D2[.,3]:>0)
W2[.,2] = D2[.,2] :- D2[.,3]*(D2[.,2]:>0) :- D2[.,4]*(D2[.,3]==0)*(D2[.,2]:>0)
W2[.,1] = D2[.,1] :- D2[.,2]*(D2[.,1]:>0) :- D2[.,3]*(D2[.,2]==0)*(D2[.,1]:>0) :-
    D2[.,4]*(D2[.,3]==0)*(D2[.,2]==0)*(D2[.,1]:>0)

// Do decision weighted utility //

V1 = W1 :* (XMAT[1,5]:^R',XMAT[1,6]:^R',XMAT[1,7]:^R',XMAT[1,8]:^R')
V1 = colsum(V1') :/ MU

V2 = W2 :* (XMAT[2,5]:^R',XMAT[2,6]:^R',XMAT[2,7]:^R',XMAT[2,8]:^R')
V2 = colsum(V2') :/ MU

```

It is a simple matter to extend this specification to handle Cumulative Prospect Theory, in which one also accounts for possible differences in the utility function used to evaluate “gains” or “losses” relative to some reference point.



### Additional References

Holt, Charles A., and Laury, Susan K., “Risk Aversion and Incentive Effects,” *American Economic Review*, 92(5), December 2002, 1644-1655.

Long, J. Scott, and Freese, Jeremy, *Regression Models for Categorical Dependent Variables Using Stata* (College Station, TX: Stata Press, Second Edition, 2006).

Quiggin, John, “A Theory of Anticipated Utility,” *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.

Saha, Atanu, “Expo-Power Utility: A Flexible Form for Absolute and Relative Risk Aversion,” *American Journal of Agricultural Economics*, 75(4), November 1993, 905-913.

StataCorp, *Stata Statistical Software: Release 10* (College Station, TX: StataCorp, 2007).

## Appendix C: Evaluating the Numerical Reliability of the Estimates

It is important to clarify exactly how these MSL estimates were obtained, since they rely intrinsically on numerical simulations and approximation methods. We initially obtained estimates by using R=50 Halton draws, the default Newton-Raphson algorithm, and allowing a modest random search for the initial point estimates; convergence was rapid, and numerically robust.<sup>19</sup> These estimates were then employed as starting values for estimation using R=500 Halton draws and looping over three algorithms.<sup>20</sup> All three algorithms converged to essentially the same maximum, at all significant digits.

These procedures provide us with a numerical “base camp” for the main series of numerical explorations, to examine the numerical fragility of the results to alternative starting values, alternative numbers of simulated likelihoods R, and alternative maximization algorithms.

Specifically, we consider a robustness check in which we (a) perturb each of the starting values  $\pm 25\%$  individually, and re-start the estimation; (b) randomly vary the number of Halton replications R between 25 and 1000; and (c) employ all three of the possible optimization algorithms. The third step entail 3 maximizations, we take 5 random drawings of R per set of starting values in the second step, and we undertake the first step 100 times, for a grand total of 1,500 maximizations. The range of values for R spans the bulk of the range suggested by Hensher and Greene [2003; p.154], who suggest that “the best test [for numerical stability] however is to estimate models over a range of draws (e.g., 25, 50, 100, 250, 500, 1,000 and 2,000 draws). Confirmation of stability/precision *for each and every model* is very important.”

The results can be conveniently evaluated using direct visual displays of the results, and some minimal data smoothing to guide the eye. The visual displays show the relationship between the estimate of interest and the number of Halton draws, using a scatterplot of the raw data and a locally

---

<sup>19</sup> Specifically, since directional vectors are calculated numerically and not analytically, no singularities or near-singularities were reported.

<sup>20</sup> Specifically, the *Stata* implementation of the Newton-Raphson algorithm, the Davidson-Fletcher-Powell algorithm, and the Broyden-Fletcher-Goldfarb-Shanno algorithm, documented in Gould, Pitblado and Sribney [2006; §1.3].

weighted regression (LOWESS) specification. This flexible specification is intended solely for descriptive purposes, and can be complemented by a fixed-effects panel regression to allow us to make for formal statements about statistically significant effects.

In this panel regression we treat the starting values as fixed effects, and regress the estimate of interest on a linear (or cubic) function of the number of Halton draws and binary dummies for the algorithms. The variables of interest are the log-likelihood value (Figure C1) and the point estimates and standard errors of the two hyper-parameters (Figure C3). Monte Carlo numerical simulations of this kind can generate outliers, and not just in one dimension, so we employ multivariate outlier detection procedures due to Hadi [1992][1994] and Gould and Hadi [1993] to remove them. Detailed estimation results for the log-likelihood are illustrated in Table C1. The first panel regression looks at the linear effect of the number of Halton draws, and the second panel regression examines the possibility of a non-linear effect by taking higher-order terms. In each case we see that the solution algorithm has no significant effect on the calculated log-likelihood, but the number of Halton draws does. The same qualitative regression results obtain for the point estimates and standard errors of the two hyper-parameters, and are not shown here.

Figure C1 shows that the precision of estimation of the log-likelihood improves monotonically with the number of Halton replications, which is to be expected under the maintained specification hypothesis that the underlying model is the true model for the population data-generating process. Of course, that specification hypothesis might be invalid, in which case the monotonic improvement we see in Figure C1 need not occur. The size of the uncertainty over the log-likelihood for Halton draws less than 400 is large enough to potentially influence the parameter estimates significantly, and this improvement becomes apparent only at relatively high levels of Halton draws. When we say “relatively high,” we have in mind the default values of 50 or 100 that are commonly employed. The top left panel of Figure C2 suggests that there is indeed a significant

effect on the point estimate of the mean and standard deviation hyper-parameters, even if their standard errors are relatively precisely estimated.<sup>21</sup>

### **Additional References**

- Gould, William, and Hadi, Ali S., “Identifying Multivariate Outliers,” in J. Hilbe (ed.), *Stata Technical Bulletin Reprints: Volume 2* (Santa Monica, CA: Computing Resource Center, 1993).
- Gould, William; Pitblado, Jeffrey, and Sribney, William, *Maximum Likelihood Estimation With Stata* (College Station, TX: Stata Press, Third Edition, 2006).
- Hadi, Ali S., “Identifying Multiple Outliers in Multivariate Data,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(3), 1992, 761-771.
- Hadi, Ali S., “A Modification of a Method for the Detection of Outliers in Multivariate Samples,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2), 1994, 393-396.

---

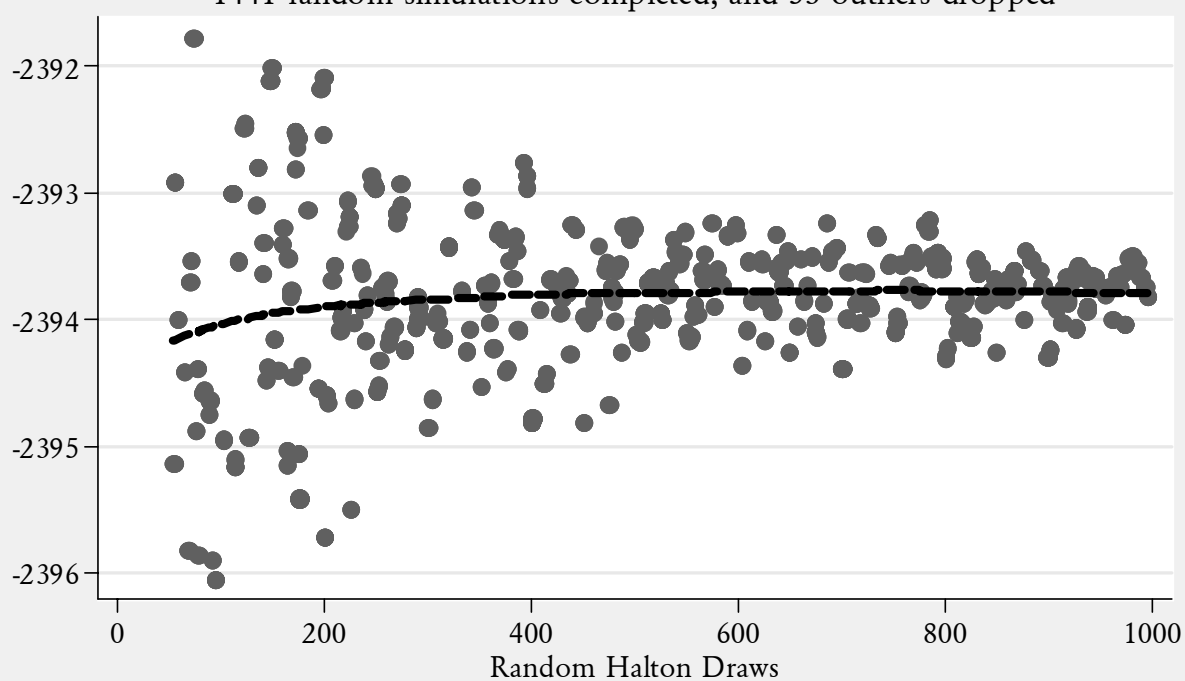
<sup>21</sup> The vertical axes in Figure C2 differ from panel to panel, and should be consulted to draw the correct inferences from comparing quantitative magnitudes.

Figure C1: Effect of Random Halton Draws on Log-likelihood

Specification with One Random Coefficient

Data smoothed with locally weighted regression (LOWESS)

1441 random simulations completed, and 53 outliers dropped

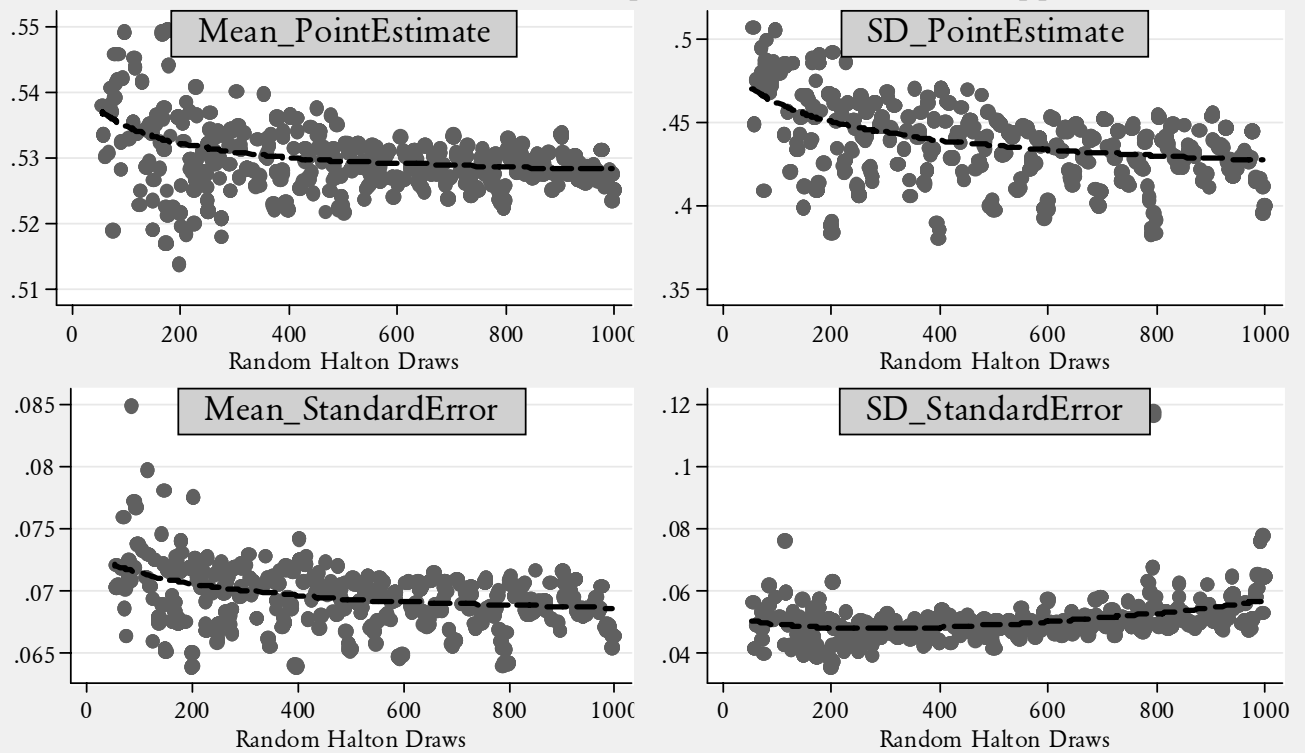


## Figure C2: Effect of Random Halton Draws on Estimates

Specification with One Random Coefficient

Data smoothed with locally weighted regression (LOWESS)

1441 random simulations completed, and 53 outliers dropped



**Table C1: Panel Regressions of Simulation Results**

```

Fixed-effects (within) regression      Number of obs      =      1388
Group variable: start_set              Number of groups    =       10

R-sq:  within  = 0.0086                Obs per group: min =       123
       between = 0.0010                avg      =      138.8
       overall  = 0.0082                max      =       150

corr(u_i, Xb)  = -0.0266                F(3,1375)          =       3.96
                                           Prob > F           =      0.0080

```

ll_	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
haltons	.0001999	.0000583	3.43	0.001	.0000856	.0003143
dfp	-.0117298	.0385365	-0.30	0.761	-.0873265	.0638669
bfgs	-.0117298	.0385365	-0.30	0.761	-.0873265	.0638669
_cons	-2393.913	.0407378	-5.9e+04	0.000	-2393.993	-2393.833
sigma_u	.05706127					
sigma_e	.58920199					
rho	.00929181	(fraction of variance due to u_i)				

```

F test that all u_i=0:      F(9, 1375) =      1.28      Prob > F = 0.2451

```

```

Fixed-effects (within) regression      Number of obs      =      1388
Group variable: start_set              Number of groups    =       10

R-sq:  within  = 0.0221                Obs per group: min =       123
       between = 0.0232                avg      =      138.8
       overall  = 0.0211                max      =       150

corr(u_i, Xb)  = -0.0432                F(5,1373)          =       6.22
                                           Prob > F           =      0.0000

```

ll_	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
haltons	.0031413	.0007655	4.10	0.000	.0016397	.004643
haltons2	-5.50e-06	1.63e-06	-3.38	0.001	-8.69e-06	-2.31e-06
haltons3	2.96e-09	1.01e-09	2.93	0.003	9.77e-10	4.95e-09
dfp	-.0098586	.038302	-0.26	0.797	-.0849954	.0652783
bfgs	-.0098586	.038302	-0.26	0.797	-.0849955	.0652782
_cons	-2394.311	.1028933	-2.3e+04	0.000	-2394.513	-2394.109
sigma_u	.05903585					
sigma_e	.58557523					
rho	.01006177	(fraction of variance due to u_i)				

```

F test that all u_i=0:      F(9, 1373) =      1.38      Prob > F = 0.1909

```

```

. test haltons haltons2 haltons3

```

```

      F( 2, 1373) =    13.06
      Prob > F =    0.0000

```