A Bayesian Approach to Characterizing Heterogeneity of Rank-Dependent Expected Utility Models of Lottery Choices

by

Dale O. Stahl
Malcolm Forsman Centennial Professor
Department of Economics
University of Texas at Austin
stahl@eco.utexas.edu

November 2016

ABSTRACT

A stylized fact from laboratory experiments is that there is much heterogeneity in human behavior. We illustrate a Bayesian approach to characterizing this heterogeneity, and apply it to laboratory data on lottery choices and the Rank-Dependent Expected Utility (RDEU) model. In addition, we define the concept of *behaviorally distinguishable* parameter vectors, and use the Bayesian posterior of the RDEU parameters to say what percentage of the population lies in meaningful regions. For instance, we find that of the subpopulation that is behaviorally distinguishable from Level-0 behavior, 84% is not behaviorally distinguishable from the Expected Utility model.

J.E.L. Classifications: C11, C12, D81

1. Introduction.

A stylized fact from laboratory experiments is that there is much heterogeneity in the subject population. How to characterize that heterogeneity is an active research area among experimentalists and econometricians. The approaches include individual parameter estimation, mixture models of different types, random coefficient models, and Bayesian methods.[1] It is not the intention of this paper to explore all the methods, but rather to illustrate how the Bayesian method can be used to characterize the heterogeneity in the population and to test models of lottery choice. In particular, we apply the Bayesian method to one of the best datasets from laboratory experiments on lottery choices: Hey and Orme (1994; hereafter HO).

The HO dataset contains 100 unique choice tasks.[2] Each task was a choice between two lotteries with three prizes drawn from the set {0£, 10£, 20£, 30£}. After all choices were completed, one task was randomly selected and the lottery the subject chose was carried out to determine monetary payoffs. On average, the difference in the expected monetary value of the two lotteries was about 5% of 30£ = 1.5£, so the expected monetary incentive for each choice task was 1.5£/100 = 0.015£ ≈ $0.02. To each decision theory, the authors appended a probit-like stochastic error specification, and computed maximum likelihood estimates of the model and error parameters for each of 80 subjects.[3] [4]

[1] E.g. see Hey and Orme (1994), Harrison and Rutström (2008, 2009), Wilcox (2008, 2011), Fox, et al. (2011), and Stahl (2014).

[2] These 100 tasks were presented to the same subjects again one week later. We do not consider that data here because the test that the same model parameters that best fit the first 100 choices are the same as those that best fit the second 100 choices fails. Possible explanations for this finding are (i) that learning took place between the sessions, (ii) preferences changed due to a change in external (and unobserved) circumstances, and (iii) the subjects did not have stable preferences. Therefore, we focus our attention on the first 100 choice tasks.

[3] Loomes and Sugden (1998) is a similar study as Hey and Orme (1994), except that their analysis of the data is based on non-parametric tests involving the number of "reversals" and violations of dominance. Harrison and Rutström (2009) replicate HO and also run a similar experiment using 30 unique tasks. Bruhin,et al. (2010) also explore heterogeneity, but they elicit certainty equivalents, so the task is arguably different from binary choices as in the other studies.

[4] Wilcox (2008, 2011) uses the entire HO data to carefully study alternative stochastic specifications and his "contextual utility" model. Briefly, contextual utility essentially rescales the payoffs for each of the choice tasks to a [0, 1] scale based on the minimum and maximum payoff in that choice task. He estimates a random parameter

The paper is organized as follows. Section 2 presents the encompassing econometric model used to analyze the HO data set, plus preliminary estimation results. Section 3 presents the Bayesian approach and interim results. Section 4 develops the formal concept of "behavioral distinguishability" that uses the Bayesian posterior to ascertain the proportion of the subjects who are or are not consistent with hypotheses about the parameters, such as: what proportion is consistent with maximizing Expected Utility (EU)? Section 5 asks and answers pertinent hypotheses. For example, we find that of the subpopulation that is behaviorally distinguishable from Level-0, 84.2% is not behaviorally distinguishable from EU. Section 6 concludes.

2. The Rank-Dependent Utility Model.

A convenient encompassing model is Rank-Dependent Expected Utility[5] (RDEU) [Quiggin (1982, 1993)], which nests EU and Expected Monetary Value (EMV). RDEU allows subjects to modify the rank-ordered cumulative distribution function of lotteries as follows. Let $Y \equiv \{y_0, y_1, y_2, y_3\}$ denote the set of potential outcomes of a lottery, where the outcomes are listed in rank order from worst to best. Given rank-ordered cumulative distribution F for a lottery on Y, it is assumed that the individual transforms F by applying a monotonic function H(F). From this transformation, the individual derives modified probabilities of each outcome:

$$h_0 = H(F_0),\ \ h_1 = H(F_1) - H(F_0),\ \ h_2 = H(F_2) - H(F_1),\text{ and }\ h_3 = 1 - H(F_2). \qquad (1)$$

A widely used parametric specification of the transformation function, suggested by Tversky and Kahneman (1992), is

$$H(F_j) \equiv (F_j)^{\beta}/[(F_j)^{\beta} + (1-F_j)^{\beta}]^{1/\beta}\ , \qquad (2)$$

where $\beta > 0$. It should be noted that to obtain the common S-shape, in which small probabilities are over-weighted and large probabilities are under-weighted, one needs $\beta < 1$. Obviously, $\beta = 1$

---

econometric model and finds that contextual utility fits and forecasts best. We take a complementary Bayesian approach to characterizing the heterogeneity across subjects, and a use a different out-of-sample forecasting test.

[5] This model is the same as the Cumulative Prospect model (Tversky and Kahneman, 1992) restricted to non-negative monetary outcomes.

corresponds to the identify transformation, in which case the RDEU model is equivalent to the EU model.

Given value function $v(y_j)$ for potential outcome y_j, the *rank-dependent expected utility* is

$$U(F) \equiv \sum_j v(y_j)h_j(F). \tag{3}$$

To confront the RDEU model with binary choice data (F^A *vs*. F^B), we assume a logistic choice function:

$$\text{Prob}(F^A) = exp\{\gamma U(F^A)\}/[\, exp\{\gamma U(F^A)\} + exp\{\gamma U(F^B)\}\,], \tag{4}$$

where $\gamma \geq 0$ is the precision parameter. As in EU theory, w.l.o.g. we can assign a value of 0 to the worst outcome and a value of 1 to the best outcome.[6] Accordingly, for the data we specify $v_0 \equiv v(y_0) = 0$ and $v_3 \equiv v(y_3) = 1$. This leaves two free utility parameters: $v_1 \equiv v(y_1)$ and $v_2 \equiv v(y_2)$. Hence, the empirical RDEU model entails four parameters: $(\gamma, v_1, v_2, \beta)$. Further, EMV is a special case of EU when $\beta = 1$, $v_1 = 1/3$ and $v_2 = 2/3$.

One can estimate these parameters for *each* subject in the HO data set. That approach entails (4×80=320) parameters, even without the corresponding variance-covariance matrices. Table 1 gives the population mean and standard deviation of the point estimates[7]. The last column "LL" gives the sum of the individually maximized log-likelihood values. Note that there is substantial heterogeneity across subjects in the parameter estimates for γ and β.

[6] Since we estimate one precision parameter for all choice tasks, this scale specification is not simply the assumption of affine invariance; it is also an assumption about the magnitude of "noise" implicit in the logistic function relative to the payoffs. Wilcox (2008) argues for a re-scaling for each choice task. While we agree that re-scaling may be needed for diverse choice tasks, we feel that in the context of the HO tasks, since all four payoffs were encountered many times in succession, a re-scaling for the entire set is more appropriate. To test our intuition, we estimated the Wilcox-type EU model for the HO data, and we found it fit slightly worse than a EU model with one precision parameter. This different finding may be due to our using only the first 100 tasks of HO and estimating parameters for each subject rather than a random coefficient specification.

[7] This is the square root of the variance of the point estimates across subjects; it is <u>not</u> the standard error of individual point estimates.

Table 1. Population Mean and Standard Deviation of Individual Parameter Estimates and the Aggregated LL

γ	v_1	v_2	β	LL
55.22 (35.56)	0.6690 (0.1782)	0.8371 (0.1039)	1.048 (0.736)	-2828.46

The previous comparisons involve estimates of a large number of parameters. For each individual subject, we obtain point estimates of the parameters, but no confidence interval. One could use a bootstrap procedure to obtain variance-covariance matrices for each individual, but that would be a computationally intense task and entail 12 additional parameters per subject. Further, the estimates for each subject would ignore the fact that the subjects are random draws from of a population of potential subjects and that therefore the behavior of the other subjects contains information that is relevant to each subject. In contrast, the Bayesian approach is better suited to extract information from the whole sample population. Consequently, we turn to the Bayesian approach.[8] We believe that this Bayesian approach is complementary to Wilcox (2008) and Fox, et al. (2011).

3. The Bayesian Approach.

To develop the Bayesian approach let x_i denote the choice data for subject i, and let $f(x_i \mid \theta)$ denote the probability of x_i given parameter vector θ. Given a prior g_0 on θ, by Bayes rule, the posterior on θ is

$$g(\theta \mid x_i) \equiv f(x_i \mid \theta)g_0(\theta)/\int f(x_i \mid z)g_0(z)dz. \qquad (5)$$

[8] For an example, see Stahl (2014).

However, eq(5) does not use information from the other subjects even though those subjects were randomly drawn from a common subject pool. Let N be the number of subjects in the data set. When considering subject i, it is reasonable to use as a prior, not g_0, but

$$g_i(\theta) \equiv \frac{1}{N-1}\sum_{h\neq i} g(\theta \mid x_h) \tag{6}$$

In other words, having observed N-1 subjects, $g_i(\theta)$ is the probability that the N^{th} random draw from the subject pool will have parameter vector θ. We then compute

$$\hat{g}_i(\theta \mid \underline{x}) \equiv f(x_i \mid \theta)g_i(\theta)/\int f(x_i \mid z)g_i(z)dz\,, \tag{7}$$

where $\underline{x}$ denotes the entire N-subject data set. Finally, we aggregate these posteriors to obtain

$$g^*(\theta \mid \underline{x}) \equiv \frac{1}{N}\sum_{i=1}^{N}\hat{g}_i(\theta \mid \underline{x})\,. \tag{8}$$

We can interpret $g^*(\theta \mid \underline{x})$ as the probability density that a random draw from the subject pool will have parameter vector θ. [Note that eq(8) puts equal weight on each x_i.]

When implementing this approach we specify the prior g_0 as follows. For the logit precision parameter, we specify $\gamma = 20\,ln[p/(1-p)]$ with p uniform on [0.5, 0.999]. In this formulation, p can be interpreted as the probability an option with a 5% greater value will be chosen. Since the mean payoff difference between lottery pairs in the HO data set is about 5%, this is a reasonable scaling factor.[9] (v_1, v_2) is uniform on the unit triangle such that $v_2 \geq v_1$. $ln(\beta)$ is uniform on $[-ln(3), ln(3)]$.[10] These three distributions are assumed to be independent. For computations, we use a grid of 41x41x41x21 = 1,477,341 points.

Since we cannot display a four-dimensional distribution, we present two two-dimensional marginal distributions. Figure 1 shows the marginal on $(p(\gamma), \beta)$, where

---

[9] The following graphs and results are robust to this specification of the prior on γ.

[10] 95% of the individual ML estimates of β lie in this range. Using a wider interval for the prior on β has no noticeable effect on the Bayesian posterior at the cost of more grid points.

$$p(\gamma) \equiv 1/[1 + exp(-0.05\gamma)].$$[11]

From Figure 1 we see that the distribution is concentrated around $\beta = 0.95$, and that the precision values are large enough to imply that a 5% difference in value is behaviorally significant (i.e. $p(\gamma) > 2/3$.

Figure 1. Marginal of g* on (p(γ), β).

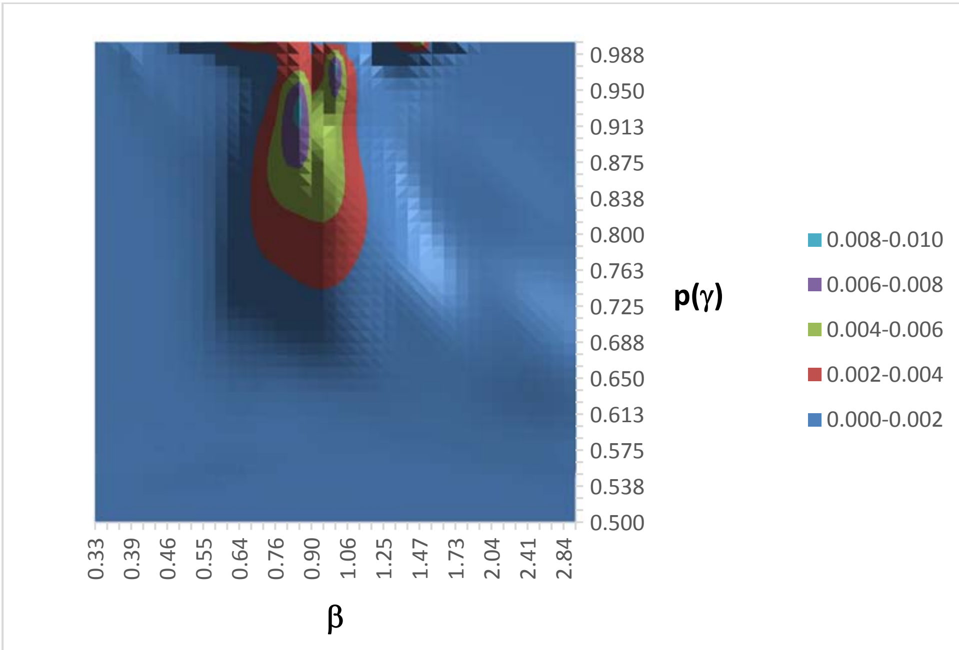
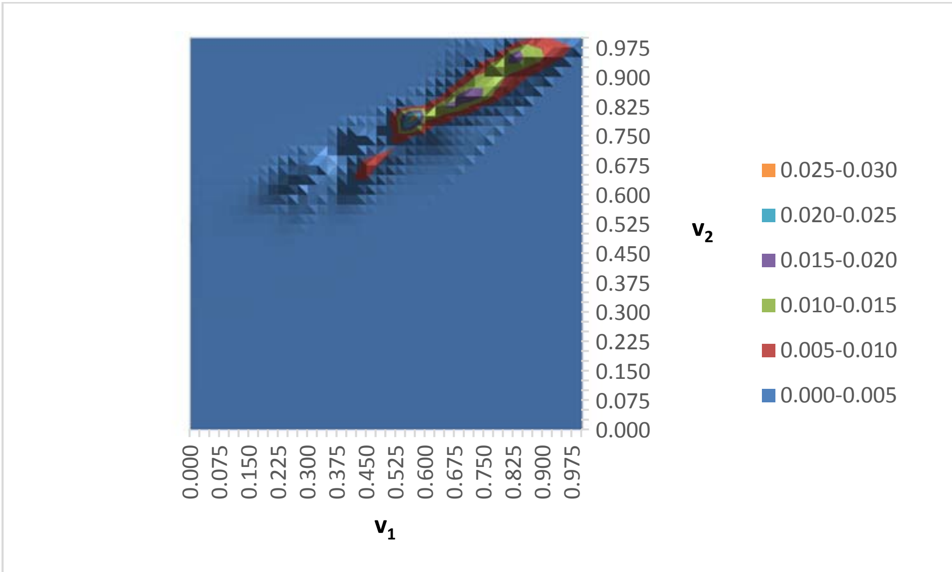


Figure 2 shows the marginal on (v_1, v_2). From Figure 2 we see that the distribution is concentrated along the line $v_2 = (v_1 + 1)/2$, which implies that utility is essentially linear above 0. We can also see a spike near the EMV point (1/3, 2/3).

[11] Thus, $p(\gamma)$ is the probability the subject will choose the option with the greater value whenever that value is 5% higher than the alternative option.

Figure 2. Marginal of g* on (v_1, v_2).



Given g*(θ | $\underline{x}$) we can compute several statistics. First, the log-likelihood of the HO data is LL(g*) = -3215.55. In contrast, the log-likelihood of the four-parameter RDEU representative-subject model is -4423.62 . Obviously, the heterogeneity implicit in g* fits the data much better than a representative-agent model.[12] Compared to -2828.46 (Table 1), the log-likelihood from the Bayesian method appears to be much worse. However, the direct comparison is inappropriate. LL(g*) is computed as if each subject were drawn independently from g*. In contrast, -2828.46 is the sum of individually computed log-likelihoods using the subject-specific estimated parameters.

The g*-weighted mean of the parameter space is $(\overline{\gamma}, \overline{v}_1, \overline{v}_2, \overline{\beta})$ = (43.3, 0.667, 0.831, 0.993); and $p(\overline{\gamma}) = 0.853$. Further, $(\overline{v}_1 + 1)/2 - \overline{v}_2 = 0.0027$, as anticipated from Figure 2. Also note that $\overline{\beta} \approx 1$, meaning that on average H(F) is the identity function. The variance-covariance matrix is

[12] One can consider this Bayesian approach as an alternative random parameter model as used by Wilcox (2008). However, in contrast to Wilcox, we assume that each subject draws from this distribution *once* and uses those parameters for all choice tasks, rather than drawing for each choice task. The latter can be viewed as a "diverse" representative agent model, while the former is a heterogeneous agent model.

	$p(\gamma)$	v_1	v_2	β
$p(\gamma)$	0.0089	0.0029	0.0028	-0.0124
v_1		0.0311	0.0172	-0.0184
v_2			0.0113	-0.0118
β				0.1418

However, these means and covariances are much less revealing when g* is multimodal.

Indeed, we find evidence for multiple modes. A grid point θ is declared a *mode* if and only if it has the highest value of g* in a 7x7x7x7 hypercube of nearest neighbors of θ in the grid. The most prominent mode is at $p(\gamma) = 0.975$, $v_1 = 0.55$, $v_2 = 0.80$, and $\beta = 1$. The next most prominent mode is at $p(\gamma) = 0.999$, $v_1 = 0.70$,$v_2 = 0.90$, and $\beta = 1.12$. The third most prominent mode is at $p(\gamma) = 0.999$, $v_1 = 0.85$, $v_2 = 0.95$, and $\beta = 0.681$. Numerous other modes exist but are best described as shallow bumps.

To test for over-fitting, we compute g* based only on the first 50 tasks in the HO data, and use this g* to predict the behavior for the second 50 tasks. We find that the log-likelihood of the latter is -1538.05. In contrast, using individual parameter estimates from just the first 50 tasks, the log-likelihood of the second 50 tasks is -1851.34. This result suggests that the approach of individual parameter estimates is more susceptible to over-fitting and less reliable than the Bayesian approach.

4. Behaviorally Indistinguishable Parameter Vectors.

The most productive use of $g^*(\theta \mid \underline{x})$ is to test hypotheses. For example, we can ask what percent of the subject pool has $\beta = 1$. The answer is 10.5%; however, this number is an artifact of the discrete grid used for computation. Assuming g* is absolutely continuous, as the grid becomes finer and finer, we would expect the percentage with $\beta = 1$ to approach 0. On the other hand, what we really want to know is the percent of the population that is *behaviorally* *indistinguishable* from EU (i.e. $\beta = 1$). The *behavior* is simply the choice data for a random subject x_i.

To assess whether this data was generated by θ or θ', we typically compute the log of the likelihood ratio (LLR): $ln[f(x_i \mid \theta)/f(x_i \mid \theta')]$. A positive LLR means x_i is more likely to have been generated by θ than θ'. However, it is well-known that likelihood-ratio tests are subject to type-I and type-II errors. To compute the expected frequency of these errors, let $X_1 \equiv \{x_i \mid ln[f(x_i \mid \theta)/f(x_i \mid \theta')] < 0\}$. If the data in fact was generated by θ, and $x_i \in X_1$, then a naïve LLR test would yield a type-I error. Similarly, if the data in fact was generated by θ' and $x_i \in X_2$ (the complement of X_1), then a naïve LLR test would yield a type-II error. Hence, the expected frequencies of type-I and type-II errors are respectively:

$$er_1 \equiv \int_{X_1} f(x_i \mid \theta)dx_i \quad \text{and} \quad er_2 \equiv \int_{X_2} f(x_i \mid \theta')dx_i \; . \tag{9}$$

If either of these error rates is too large, we might say that θ and θ' are behaviorally indistinguishable. Classical statistics suggests that a proper test statistic would have these error rates not exceed 5%.

Of course, by increasing the number of observations in x_i, we can drive these error rates lower and lower. However, practical considerations often limit the number of observations. In laboratory experiments, boredom, time limitations and budget constraints place severe upper bounds on the number of observations. The HO dataset with 100 tasks is unusually large. Moreover, to test for overfitting we would select a subset, say 50, to use for estimation, and the remaining 50 to assess parameter stability and prediction performance. Therefore, for the illustrative purposes of this paper, we use 50 as a reasonable sample size upon which to judge behavioral distinguishability. With 50 binary choices, there are 2^{50} ($\approx 10^{30}$) possible x_i vectors. For the tests we want to conduct, generating all these possible x_i vectors and computing er_1 and er_2 is obviously not feasible. Instead, we generate 1000 x_i vectors from $f(x_i \mid \theta)$ and 1000 from $f(x_i \mid \theta')$.[13] Then, er_1 is approximated by the proportion of x_i generated by $f(x_i \mid \theta)$ that lie in X_1, and er_2 is approximated by the proportion of x_i generated by $f(x_i \mid \theta')$ that lie in X_2. In summary,

---

[13] We also made these computations with only 100 simulated x_i vectors, and found virtually the same results. Therefore, we are confident that 1000 simulated x_i vectors are adequate for our purposes.

we declare θ and θ' to be *behaviorally indistinguishable* if either of the simulated type-I and type-II error rates exceed 5%, and to be *behaviorally distinguishable* if both of the simulated type-I and type-II error rates are less than or equal to 5%,

5. Questions and Answers.

The questions we are interested in answering are easily framed in terms of our *behaviorally indistinguishable* relationship on the parameters. To begin, we want to know what percent of the population is behaviorally indistinguishable from 50:50 random choices (hereafter referred to as Level-0 behavior). Since the latter entails the simple restriction that $\gamma = 0$, we can compute whether $\theta = (\gamma, u_1, u_2, \beta)$ is behaviorally indistinguishable from $(0, v_1, v_2, \beta)$, and then sum $g^*(\gamma, v_1, v_2, \beta)$ over all the grid points $(\gamma, \beta, v_1, v_2)$ that are behaviorally indistinguishable from $(0, v_1, v_2, \beta)$. The answer is 4.0%, which leaves 96.0% that is behaviorally distinguishable from Level-0. We are not interested in dissecting Level-0 behavior. Therefore, all our subsequent questions are conditional on the parameters being behaviorally distinguishable from Level-0.

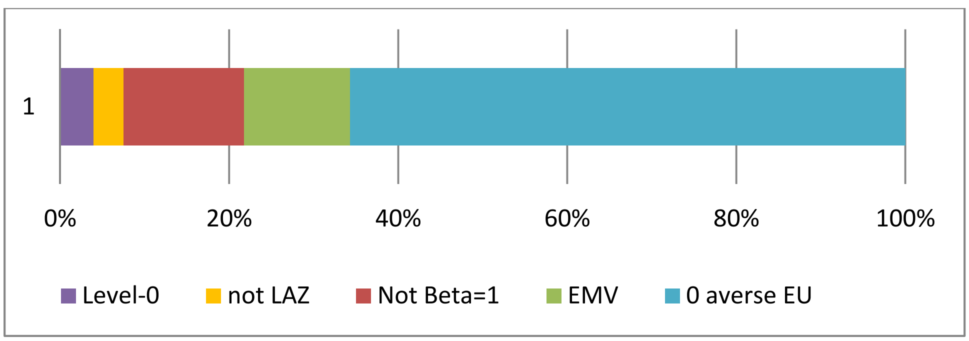
Since Figure 2 provides strong evidence that the utility function is Linear Above Zero (LAZ), our next question is what percent of the population is behaviorally distinguishable from Level-0 but behaviorally indistinguishable from LAZ? The latter criteria can be stated as: $(\gamma, v_1, v_2, \beta)$ is behaviorally indistinguishable from $(\gamma, v_1, (1+v_1)/2, \beta)$. The answer is 92.5%. Hence, of the population that is behaviorally distinguishable from Level-0, 96.4% (= 92.5/96.0) is behaviorally indistinguishable from LAZ.

Perhaps the question of most interest is what percent are behaviorally indistinguishable from EU. To answer this, we ask how much mass g* puts on the set of parameters $(\gamma, v_1, v_2, \beta)$ that are behaviorally distinguishable from Level-0 but indistinguishable from $(\gamma, v_1, (1+v_1)/2, 1)$? The answer is 78.2%. Hence, of the subpopulation that is behaviorally distinguishable from Level-0, 84.5% (= 78.2/96.0) is behaviorally indistinguishable from EU.

Our fourth and final question concerns the apparent aversion to 0 payoffs. What percent of the population are *pure EMVs* (i.e. maximize EMV with no aversion for 0 payoffs)? This additional criteria can be stated as: (γ, v_1, v_2, β) is behaviorally indistinguishable from (γ, 1/3, 2/3, 1). The answer is 12.5%. Hence, of the EU and LAZ subpopulation, 17.3% (= 12.5/78.2) have no aversion to \$0 and 82.7% are averse to \$0. Aversion to 0 is akin to loss aversion[14], and the latter is a common result in the psychology literature (e.g. Kahneman and Tversky, 1979; Erev, Ert and Yechiam, 2008).

Figure 3 conveniently gathers these results in a bar graph. The first section labelled Level-0 represents the 4.0% that are behaviorally indistinguishable from Level-0. The second section labelled "not LAZ" represents the 3.5% (=96 – 92.5) that are behaviorally distinguishable from Level-0 and LAZ. The third section labelled "Not β=1" represents the 14.3% (= 92.5 – 78.2) that are behaviorally distinguishable from Level-0 and β=1 but not from LAZ. The fourth section represents the 12.5% that are pure EMV maximizers. The fifth and final section represents the 65.7% (= 78.2 – 12.5) that EMV maximizers but with an aversion to 0.

Figure 3. Subdivisions of the Subject Population.



6. <u>Conclusions and Discussion</u>.

Our Bayesian analysis has characterized substantial heterogeneity in the subject population. On the other hand it has revealed that 78.2% of the population is behaviorally indistinguishable from EU behavior (84.2% of the subpopulation that is behaviorally distinguishable from Level-0). This finding reinforces Hey and Orme (1994)'s conclusion: "Our

[14] With prizes of {0£, 10£, 20£, 30£}, the last three are clear wins, while 0£ is not a win and could easily be interpreted by the DM as losing relative to aspirations.

study indicates that behavior can be reasonably well modelled (to what might be termed a ‘reasonable approximation’) as EU plus noise. Perhaps we should now spend some time on thinking about the noise, rather than about even more alternatives to EU.”

Another interesting finding is that the vast majority of subjects are behaviorally indistinguishable from having a linear utility function from 10£ to 30£, although a majority exhibit an aversion to a 0 payoff.

We hope this paper has demonstrated the feasibility and usefulness of Bayesian methods when confronting laboratory data, especially when addressing heterogeneous behavior. To extend our approach to models with more parameters, statistical sampling techniques can be employed to tame the curse of dimensionality.[15]

[15] E.g. see Rubinstein and Kroese (2016).

References

Bruhin, A., Fehr-Duda, H., and Epper, T. (2010). "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica,* **78**, 1375-1412.

Erev, I., Ert, E. and Yechiam, E. (2008). "Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions," *Journal of Behavioral Decision Making,* **21**, 575–597.

Fox, J., Kim, K., Ryan, S. and Bajari, P. (2011). "A Simple Estimator for the Distribution of Random Coefficients," *Quantitative Economics,* **2**, 381-418.

Grether, D. and Plott, C. (1979). "Economic Theory of Choice and Preference Reversal Phenomena," *Am. Econ. Rev.,* **69**, 623-638.

Harrison, G. W. (1989). "Theory and Misbehavior of First-Price Auctions," *Am. Econ. Rev.,* **79**, 749-762.

Harrison, G. W. (1992). "Theory and Misbehavior of First-Price Auctions: Reply," *Am. Econ. Rev.,* **82**, 1426-1443.

Harrison, G. W. and Rutström, E. (2008). "Risk Aversion in the Laboratory," In J. C. Cox and G. W. Harrison, eds., *Research in Experimental Economics,* **12**, 41-196.

Harrison, G. W. and Rutström, E. (2009). "Expected Utility and Prospect Theory: One Wedding and Decent Funeral," *Experimental Economics*, 12, 133-158.

Harrison, G. W., A. Swarthout, J. T (2014). "Experimental Payment Protocols and the Bipolar Behaviorist," *Theory and Decision,* 77, 423-438.

Hey, J. and Orme, C. (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62, 1291-1326.

Kahneman, D. and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica,* **47**, 263–291.

Loomes G., and Sugden, R. (1998). "Testing Alternative Stochastic Specifications for Risky Choice," *Economica*, **65**, 581-598.

Machina, M. (2008). "Non-expected Utility Theory," in *The New Palgrave Dictionary of Economics* (2nd edition). Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

Quiggin, J. (1982). "A Theory of Anticipated Utility", *J. of Econ. Behavior and Organization*, **3**, 323-343.

Quiggin, J. (1993). *Generalized Expected Utility Theory: the Rank-Dependent Model,* Kluwer Academic Publishers.

Rubinstein, B. and Kroese, D. (2016). *Simulation and the Monte Carlo Method*, 3rd edition, Wiley & Sons.

Stahl, D. (2014). "Heterogeneity of Ambiguity Preferences," *Review of Economics and Statistics,* **96**, 609-617.

Tversky, A., and Kahneman, D. (1992). "Cumulative Prospect Theory: An Analysis of Decision Under Uncertainty," *Journal of Risk and Uncertainty*, **5**, 297-323.

Wilcox, N. (2008). "Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison". In J. C. Cox and G. W. Harrison, eds., *Research in Experimental Economics,* **12**, 197-292.

Wilcox, N. (2011). "Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk," *Journal of Econometrics*, **162**, 89–104.