

Behavioral Econometrics for Psychologists

by

Steffen Andersen, Glenn W. Harrison, Morten Igel Lau and Elisabet E. Rutström[†]

March 2010

Forthcoming, *Journal of Economic Psychology*

Abstract. We make the case that psychologists should make wider use of econometric methods for the estimation of structural models. These methods involve the development of maximum likelihood estimates of models, where the likelihood function is tailored to the structural model. In recent years these models have been developed for a wide range of behavioral models of choice under uncertainty. We explain the components of this methodology, and illustrate with applications to major models from psychology. The goal is to build, and traverse, a constructive bridge between the modeling insights of psychology and the statistical tools of economists.

[†] Department of Economics, Copenhagen Business School, Copenhagen, Denmark (Andersen); Department of Risk Management & Insurance, and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, USA (Harrison), and Department of Economics, College of Business Administration, University of Central Florida, USA (Rutström) and Department of Economics and Finance, Durham Business School, Durham University, United Kingdom (Lau). E-mail: SA.ECO@CBS.DK, GHARRISON@GSU.EDU, M.I.LAU@DURHAM.AC.UK, and ERUTSTROM@BUS.UCF.EDU. Harrison and Rutström thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/HSD 0527675 and NSF/SES 0616746, and we all thank the Danish Social Science Research Council for research support under project #24-02-0124. An earlier version was presented by Harrison as an invited lecture at the *Annual Conference of the French Economic Association: Behavioral Economics and Experimental Economics*, Lyon, May 23-25, 2007, and we are grateful for the invitation, hospitality and comments received there. We are also grateful for comments from George Judge, Daniel Read and a referee.

Table of Contents

| | |
|--|-------|
| 1. Elements of the Estimation of Structural Models | -2- |
| 1.1 Estimation of a Structural Model Assuming EUT | -2- |
| 1.2 Stochastic Errors | -6- |
| 2. Probability Weighting and Rank-Dependent Utility | -9- |
| 3. Loss Aversion and Sign-Dependent Utility | -12- |
| 3.1 Original Prospect Theory | -12- |
| 3.2 Cumulative Prospect Theory | -14- |
| 3.3 Will the True Reference Point Please Stand Up? | -15- |
| 4. Mixture Models and Multiple Decision Processes | -18- |
| 4.1 Recognizing Multiple Decision Processes | -18- |
| 4.2 Implications for the Interpretation of Process Data | -20- |
| 4.3 Comparing Latent Process Models | -20- |
| 5. Dual Criterion Models from Psychology | -23- |
| 5.1 The “SP” Criterion | -25- |
| 5.2 The “A” Criterion | -26- |
| 5.3 Combining the Two Criterion | -27- |
| 5.4 Applications | -29- |
| 6. The Priority Heuristic | -32- |
| 7. Conclusion | -38- |
| References | -45- |
| Appendix A: Experimental Procedures for Binary Lottery Choice Task | -A1- |
| Appendix B: Estimation Using Maximum Likelihood | -A3- |
| B1. Estimating a CRRA Utility Function | -A4- |
| B2. Loss Aversion and Probability Weighting | -A9- |
| B3. Adding Stochastic Errors | -A11- |
| B4. Extensions | -A13- |
| Appendix C: Instructions for Laboratory Experiments | -A15- |
| C1. Baseline Instructions | -A15- |
| C2. Additional Instructions for UK Version | -A22- |

Economists tend not to take the full range of theories from psychology as seriously as they should. Psychologists have much more to offer to economists than the limited array of models or *ad hoc* insights that have been adopted by behavioral economists. One simple reason for this lack of communication is that psychologists tend to “estimate” their models, and test them, in a naive fashion that makes it hard for economists to evaluate the broader explanatory power of those models. Another reason is that psychologists tend to think in terms of the process of decision-making, rather than the characterization of the choice itself, and it has been hard to see how such models could be estimated in the same way as standard models from economics. We propose that psychologists use the maximum likelihood estimation of structural models to address these barriers to trade between the two disciplines (or that economists use these methods to evaluate models from psychology).

Recent developments in “behavioral econometrics” allow much richer specifications of traditional and non-traditional models of behavior. It is possible to jointly estimate parameters of complete structural models, rather than using one experiment to pin down one parameter, another experiment to pin down another, and losing track of the explanatory power and sampling errors of the whole system. It is also possible to see the maximum likelihood evaluator as an intellectual device to write out the process in as detailed a fashion as desired, rather than relying on pre-existing estimation routines to shoe-horn the model into. The mainstream models can also be seen as process models in this light, even if they do not need to be interpreted that way in economics.

In section 1 we review the basic elements of structural modeling of choice under uncertainty, using expected utility theory from mainstream economics to illustrate and provide a baseline model. In section 2 we illustrate how one of the most important insights from psychology (Edwards [1962]), the possibility of probability weighting, can be incorporated. In section 3 we demonstrate the effects of including one of the other major insights from psychology (Kahneman and Tversky [1979]), the possibility of sign dependence in utility evaluation. In particular, we demonstrate how circularity in the use of priors about the true reference point can dramatically affect the empirical inferences one

might make about the prevalence of loss aversion. The introduction of alternative structural models leads to a discussion of how one should view the implied hypothesis testing problem. We advocate a mixture specification in section 4, in which one allows for multiple latent data-generating processes, and then uses the data to identify which process applies in which task domain and for which subjects. We then examine in section 5 how the tools of behavioral econometrics can be applied to the neglected model from psychology proposed by Lopes [1995]. Her model represents a novel way to think about rank-dependent and sign-dependent choices jointly, and directly complements literature in economics. Finally, in section 6 we examine the statistical basis of the claims of Brandstatter, Gigerenzer and Hertwig [2006] that their Priority Heuristic dramatically outperforms other models of choice under uncertainty from economics and psychology.

We are not saying that psychologists are ignorant about the value or methods of maximum likelihood and structural models, that every behavioral economist uses these methods, or indeed that they are needed for every empirical issue that arises between economists and psychologists. Instead, we are arguing that many needless debates can be efficiently avoided if we share a common statistical language for communication. The use of experiments themselves provides a critical building block in developing that common language: if we can resolve differences in procedures then the experimental data itself provides an objective basis for debates over interpretation to be meaningfully joined.¹

1. Elements of the Estimation of Structural Models

1.1 Estimation of a Structural Model Assuming EUT

Assume for the moment that utility of income is defined by

$$U(x) = x^r \tag{1}$$

where x is the lottery prize and r is a parameter to be estimated. For $r=0$ assume $U(x)=\ln(x)$ if needed. Thus $1-r$ is the coefficient of Constant Relative Risk Aversion (CRRA): $r=1$ corresponds to risk neutrality, $r>1$ to risk loving, and $r<1$ to risk aversion. Let there be K possible outcomes in a

¹ Hertwig and Ortmann [2001][2005] evaluate the differences systematically, and in a balanced manner.

lottery. Under EUT the probabilities for each outcome k , p_k , are those that are induced by the experimenter, so expected utility is simply the probability weighted utility of each outcome in each lottery i :

$$EU_i = \sum_{k=1,K} [p_k \times u_k]. \quad (2)$$

The EU for each lottery pair is calculated for a candidate estimate of r , and the index

$$\nabla EU = EU_R - EU_L \quad (3)$$

calculated, where EU_L is the “left” lottery and EU_R is the “right” lottery. This latent index, based on latent preferences, is then linked to the observed choices using a cumulative normal distribution function $\Phi(\cdot)$. This “probit” function takes any argument between $\pm\infty$ and transforms it into a number between 0 and 1. Figure 1 illustrates the standard cumulative normal distribution function where $\sigma=1$. The agent chooses lottery R if $\nabla EU + \epsilon > 0$, where ϵ is a normally distributed error term with mean zero and variance σ^2 . Thus we have the probit link function, showing the probability that R is chosen, as

$$\begin{aligned} \text{prob}(\text{choose lottery R}) &= P(\nabla EU + \epsilon > 0) = \\ &P(\epsilon/\sigma > -\nabla EU/\sigma) = \Phi(\nabla EU/\sigma) \end{aligned} \quad (4)$$

The logistic function is very similar, as illustrated in Figure 1, and leads instead to the “logit” specification.²

Even though Figure 1 is common in econometrics texts, it is worth noting explicitly and understanding. It forms the critical statistical link between observed binary choices, the latent structure generating the index y^* , and the probability of that index y^* being observed. In our applications y^* refers to some function, such as (3), of the EU of two lotteries; or, later, the prospective utility of two lotteries. The index defined by (3) is linked to the observed choices by specifying that the R lottery is chosen when $\nabla EU + \epsilon > 0$, which is implied by (4).

Thus the likelihood of the observed responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of r given the above statistical specification and

² One often sees formally identical specifications in the psychology literature. For example, Birnbaum and Chavez [1997; p.187, eq. (14)] specify that $\text{prob}(\text{choose lottery R}) = \Phi(\hat{\alpha} \times \nabla EU)$ and estimate $\hat{\alpha}$.

the observed choices. The “statistical specification” here includes assuming some functional form for the cumulative density function (CDF), such as one of the two shown in Figure 1. If we ignore responses that reflect indifference³ the conditional log-likelihood would be

$$\ln L(r; y, \mathbf{X}) = \sum_i [(\ln \Phi(\nabla EU)) \times \mathbf{I}(y_i = 1) + (\ln(1 - \Phi(\nabla EU))) \times \mathbf{I}(y_i = -1)] \quad (5)$$

where $\mathbf{I}(\cdot)$ is the indicator function, $y_i = 1(-1)$ denotes the choice of the Option R (L) lottery in risk aversion task i , and \mathbf{X} is a vector of individual characteristics reflecting age, sex, race, and so on.⁴

The parameter r is defined as a linear function of the characteristics in vector \mathbf{X} .

The latent index (3) could have been written in a ratio form:

$$\nabla EU = EU_R / (EU_R + EU_L) \quad (3')$$

and then the latent index would already be in the form of a probability between 0 and 1, so we would not need to take the probit or logit transformation. This specification has been used, with some modifications to include stochastic errors, in Holt and Laury [2002].

Appendix A reviews experimental procedures for some canonical binary lottery choice tasks we will use to illustrate many of the models considered here. These data amount to a replication of the classic experiments of Hey and Orme [1994], with extensions to collect individual demographic characteristics and to present subjects with some prizes framed as losses. Details of the experiments are reported in Harrison and Rutström [2008][2009]. Subjects were recruited from the undergraduate and graduate student population of the University of Central Florida in late 2003 and throughout 2004. Each subject made 60 lottery choices and was paid for 3 of these, drawn at random. A total of 158 subjects made choices. Some of these had prizes of \$0, \$5, \$10 and \$15 in what we refer to as the gain frame ($N=63$). Some had prizes framed as losses of \$15, \$10, \$5 and \$0 relative to an

³ Relatively few subjects use this option. The extension to handling it in models of this kind is discussed in Harrison and Rutström [2008; §2.2].

⁴ The likelihood function is sometimes written in the formally sloppy form $\ln L(r; y, \mathbf{X}) = \sum_i [(\ln \Phi(\nabla EU) \mid y_i = 1) + (\ln(1 - \Phi(\nabla EU)) \mid y_i = -1)]$, where it is understood that the log-likelihood is 0 when the conditional in each of the terms on the right hand side is not true. This form is popular with programmers, who are accustomed to such default assignments when conditionals evaluate to be false. A more succinct form, popular with econometricians, exploits the Boolean manner in which the binary variable y is normally defined, with $y=1$ for one choice and $y=-1$ for the other choice, to write the likelihood as $\ln L(r; y, \mathbf{X}) = \sum_i [\ln \Phi(\nabla EU \times y_i)]$. We prefer not to use this compact form, since it relies on the variable y being defined a certain way, and does not allow students to easily see how to generalize the likelihood, such as when indifference is a third choice. In any event, these are all seen to refer to the same overall log-likelihood.

endowment of \$15, ending up with the same final prize outcomes as the gain frame (N=58). Finally, some subjects had an endowment of \$8, and the prizes were transformed to be -\$8, -\$3, \$3 and \$8, generating final outcomes inclusive of the endowment of \$0, \$5, \$11 and \$16 (N=37).

Appendix B reviews procedures and syntax from the popular statistical package *Stata* that can be used to estimate structural models of this kind, as well as more complex models discussed later.⁵ The goal is to illustrate how to write explicit maximum likelihood (ML) routines that are specific to different structural choice models. It is a simple matter to correct for stratified survey responses, multiple responses from the same subject (“clustering”),⁶ or heteroskedasticity, as needed, and those procedures are discussed in Appendix B.

Panel A of Table 1 shows maximum likelihood estimates obtained with this simple specification (equations 1- 5) with $\sigma=1$. The coefficient r is estimated to be 0.777, with a 95% confidence interval between 0.729 and 0.825. This indicates modest degrees of risk aversion, consistent with vast amounts of experimental evidence for samples of this kind.

Extensions of the basic model are easy to implement, and this is the major attraction of this approach to the estimation of structural models. For example, one can easily extend the functional forms of utility to allow for varying degrees of relative risk aversion (RRA). Consider, as one important example, the Expo-Power (EP) utility function proposed by Saha [1993]. Following Holt and Laury [2002], the EP function is defined as

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

where α and r are parameters to be estimated. RRA is then $r + \alpha(1-r)y^{1-r}$, so RRA varies with income if $\alpha \neq 0$. This function nests CRRA (as $\alpha \rightarrow 0$) and CARA (as $r \rightarrow 0$).

⁵ Appendix B is available in the working paper version available from the authors.

⁶ Clustering commonly arises in national field surveys from the fact that physically proximate households are often sampled to save time and money, but it can also arise from more homely sampling procedures. For example, Williams [2000; p.645] notes that it could arise from dental studies that “collect data on each tooth surface for each of several teeth from a set of patients” or “repeated measurements or recurrent events observed on the same person.” The procedures for allowing for clustering allow heteroskedasticity between and within clusters, as well as autocorrelation within clusters. They are closely related to the “generalized estimating equations” approach to panel estimation in epidemiology (see Liang and Zeger [1986]), and generalize the “robust standard errors” approach popular in econometrics (see Rogers [1993]). Wooldridge [2003] reviews some issues in the use of clustering for panel effects, noting that significant inferential problems may arise with small numbers of panels.

It is also a simple matter to generalize this ML analysis to allow the core parameter r to be a linear function of observable characteristics of the individual or task. For example, assume that we collected information on the sex of the subject, and coded this as a binary dummy variable called Female. In this case we extend the model to be $r = r_0 + r_1 \times \text{Female}$, where r_0 and r_1 are now the parameters to be estimated. In effect the prior model was to assume $r = r_0$ and just estimate r_0 . This extension significantly enhances the attraction of ML estimation of structural models, particularly for responses pooled over different subjects, since one can condition estimates on observable characteristics of the task or subject. We illustrate the richness of this extension later. For now, we estimate $r_0=0.83$ and $r_1= -0.12$, with standard errors of 0.029 and 0.050 respectively. So there is some evidence of a sex effect, with women exhibiting slightly greater risk aversion. Of course, this specification does not control for other variables that might be correlated with the effect of sex.

1.2 Stochastic Errors

An important extension of the core model is to allow for subjects to make some errors. The notion of error is one that has already been encountered in the form of the statistical assumption that the probability of choosing a lottery is not 1 when the EU of that lottery exceeds the EU of the other lottery. This assumption is clear in the use of a link function between the latent index ∇EU and the probability of picking one or other lottery; in the case of the normal CDF, this link function is $\Phi(\nabla EU/\sigma)$ and is displayed in Figure 1 for $\sigma=1$. If there were no sampling errors ($\sigma=0$), and the subject exhibited no errors from the perspective of EUT, this function would be a step function in Figure 1: zero for all values of $y^*<0$, anywhere between 0 and 1 for $y^*=0$, and 1 for all values of $y^*>0$. By varying the shape of the link function in Figure 1, one can informally imagine subjects that are more sensitive to a given difference in the index ∇EU and subjects that are not so sensitive. Of course, such informal intuition is not strictly valid, since we can choose any scaling of utility for a given subject, but it is suggestive of the motivation for allowing for errors, and why we might want them to vary across subjects or task domains.

Consider the error specification used by Holt and Laury [2002], originally due to Luce [1959], and popularized by Becker, DeGroot and Marschak [1963]. The EU for each lottery pair is calculated for candidate estimates of r , as explained above, and the ratio

$$\nabla EU = EU_R^{1/\mu} / (EU_L^{1/\mu} + EU_R^{1/\mu}) \quad (3'')$$

calculated, where μ is a structural “noise parameter” used to allow some errors from the perspective of the deterministic EUT model. The index ∇EU is in the form of a cumulative probability distribution function defined over differences in the EU of the two lotteries and the noise parameter μ . Thus, as $\mu \rightarrow 0$ this specification collapses to the deterministic choice EUT model, where the choice is strictly determined by the EU of the two lotteries; but as μ gets larger and larger the choice essentially becomes random. When $\mu=1$ this specification collapses to (3'), where the probability of picking one lottery is given by the ratio of the EU of one lottery to the sum of the EU of both lotteries. Thus μ can be viewed as a parameter that flattens out the link functions in Figure 1 as it gets larger. This is just one of several different types of error story that could be used, and Wilcox [2008] provides a masterful review of the implications of the alternatives.⁷

There is one other important error specification, due originally to Fechner [1860] and popularized by Becker, DeGroot and Marschak [1963] and Hey and Orme [1994]. This error specification posits the latent index

$$\nabla EU = (EU_R - EU_L)/\mu \quad (3''')$$

instead of (3), (3') or (3''), and allows μ to differ from 1. When (3''') is combined with the probit specification (4), μ is formally identical to σ . The conventional interpretation is that σ refers to sampling errors, and that the structural parameter μ refers to behavioral errors. In our analyses we

⁷ See Harless and Camerer [1994], Hey and Orme [1994] and Loomes and Sugden [1995] for the first wave of empirical studies including some formal stochastic specification in the version of EUT tested. There are several species of “errors” in use, reviewed by Hey [1995][2002], Loomes and Sugden [1995], Ballinger and Wilcox [1997], and Loomes, Moffatt and Sugden [2002]. Some place the error at the final choice between one lottery or the other after the subject has decided deterministically which one has the higher expected utility; some place the error earlier, on the comparison of preferences leading to the choice; and some place the error even earlier, on the determination of the expected utility of each lottery. Within psychology, Birnbaum [2004b; p.57-63] discusses the implications of these and other error specifications for observed choice patterns, in the spirit of Harless and Camerer [1994]. However, he does not integrate them into estimation of the structural models of choice under uncertainty he is testing, in the spirit of Hey and Orme [1994]. However, the model estimated in Birnbaum and Chavez [1997; p. 187, eq.(14)] can be viewed as formally equivalent to a Fechner specification, as noted earlier.

default to the use of the Fechner specification, but recognize that we need to learn a great deal more about how these stochastic error specifications interact with substantive inferences (e.g., Loomes [2005], Wilcox [2008][2009], Harrison and Rutström [2008; §2.3]).

Panel B of Table 1 illustrates the effect of incorporating a Fechner error story into the basic EUT specification of Panel A. In this case there is virtually no change in the point estimate of risk attitudes, but a slight widening of the confidence interval.

Panel C illustrates the effects of allowing for observable individual characteristics in this structural model. The core coefficients r and μ are each specified as a linear function of several characteristics. The heterogeneity of the error specification μ is akin to allowing for heteroskedasticity, but it is important not to confuse the structural error parameter μ from the sampling errors associated with parameter estimates. We observe that the effect of sex remains statistically significant, now that we include some potential confounds. We also see an effect from being Hispanic, associated with an increase in risk aversion. Although barely statistically significant, with a p -value of 0.103, every extra year in age is associated with the subject being less risk averse by 0.030 in CRRA terms. Figure 2 displays the distribution of predicted risk attitudes from the model estimated in Panel C of Table 1. The average of this distribution is 0.76, close to the point estimate from Panel B, of course.

It is a simple matter to specify different choice models, and this is perhaps the main advantage of estimating structural models since non-EUT choice models tend to add more structural parameters. We now consider extensions to such non-EUT models.

2. Probability Weighting and Rank-Dependent Utility

One route of departure from EUT has been to allow preferences to depend on the rank of the final outcome through probability weighting. The idea that one could use non-linear transformations of the probabilities as a lottery when weighting outcomes, *instead* of non-linear transformations of the outcome into utility, was most sharply presented by Yaari [1987]. To illustrate

the point clearly, he assumed a linear utility function, in effect ruling out any risk aversion or risk seeking from the shape of the utility function *per se*. Instead, concave (convex) probability weighting functions would imply risk seeking (risk aversion).⁸ It was possible for a given decision maker to have a probability weighting function with both concave and convex components, and the conventional wisdom held that it was concave for smaller probabilities and convex for larger probabilities.

The idea of rank-dependent preferences for choice over lotteries had two important precursors.⁹ In economics Quiggin [1982] had formally presented the general case in which one allowed for subjective probability weighting in a rank-dependent manner and allowed non-linear utility functions. This branch of the family tree of choice models has become known as Rank-Dependent Utility (RDU). The Yaari [1987] model can be seen as a pedagogically important special case, and can be called Rank-Dependent Expected Value (RDEV). The other precursor, in psychology, is Lopes [1984]. Her concern was motivated by clear preferences that experimental subjects exhibited for lotteries with the same expected value but alternative shapes of probabilities, as well as the verbal protocols those subjects provided as a possible indicator of their latent decision processes.

Formally, to calculate decision weights under RDU one replaces expected utility

$$EU_i = \sum_{k=1, K} [p_k \times u_k] \tag{2}$$

with RDU

$$RDU_i = \sum_{k=1, K} [w_k \times u_k] \tag{2'}$$

where

$$w_i = \omega(p_i + \dots + p_n) - \omega(p_{i+1} + \dots + p_n) \tag{6a}$$

⁸ Camerer [2005; p.130] provides a useful reminder that “Any economics teacher who uses the St. Petersburg paradox as a “proof” that utility is concave (and gives students a low grade for not agreeing) is confusing the sufficiency of an explanation for its necessity.”

⁹ Of course, many others recognized the basic point that the distribution of outcomes mattered for choice in some holistic sense. Allais [1979; p.54] was quite clear about this, in a translation of his original 1952 article in French. In psychology, Birnbaum, Coffey, Mellors and Weiss [1992] provide extensive cites to “configural-weight” models that have a close relationship to rank-dependent specifications. Similarly, it is easy to find citations to kindred work in psychology in the 1960's and 1970's by Lichtenstein, Coombs, Payne and Birnbaum *inter alia*.

for $i=1, \dots, n-1$, and

$$w_i = \omega(p_i) \tag{6b}$$

for $i=n$, the subscript indicates outcomes ranked from worst to best, and where $\omega(p)$ is some probability weighting function.

Picking the right probability weighting function is obviously important for RDU specifications. A weighting function proposed by Tversky and Kahneman [1992] has been widely used. It is assumed to have well-behaved endpoints such that $\omega(0)=0$ and $\omega(1)=1$ and to imply weights

$$\omega(p) = p^\gamma / [p^\gamma + (1-p)^\gamma]^{1/\gamma} \tag{7}$$

for $0 < p < 1$. The normal assumption, backed by a substantial amount of evidence reviewed by Gonzalez and Wu [1999], is that $0 < \gamma < 1$. This gives the weighting function an “inverse S-shape,” characterized by a concave section signifying the overweighting of small probabilities up to a crossover-point where $\omega(p)=p$, beyond which there is then a convex section signifying underweighting. Under the RDU assumption about how these *probability* weights get converted into *decision* weights, $\gamma < 1$ implies *overweighting* of extreme outcomes. Thus the probability associated with an outcome does not directly inform one about the decision weight of that outcome. If $\gamma > 1$ the function takes the less conventional “S-shape,” with convexity for smaller probabilities and concavity for larger probabilities.¹⁰ Under RDU $\gamma > 1$ implies *underweighting* of extreme outcomes.

We again assume the CRRA functional form

$$U(x) = x^\rho \tag{1''}$$

for utility. The parameter ρ determines the concavity of the utility function but is no longer the sole determinant of risk attitudes since probabilities are also transformed. The remainder of the econometric specification is the same as for the EUT model with Fechner error μ , generating

¹⁰ There are some well-known limitations of the probability weighting function (7). It does not allow independent specification of location and curvature; it has a crossover-point at $p=1/e=0.37$ for $\gamma < 1$ and at $p=1-0.37=0.63$ for $\gamma > 1$; and it is not increasing in p for small values of γ . Prelec [1998] and Rieger and Wang [2006] offer two-parameter probability weighting functions that exhibits more flexibility than (7), but for our expository purposes the standard probability weighting function is adequate. Harrison, Humphrey and Verschoor [2010] explore the applied use of these more flexible functional forms.

$$\nabla RDU = (RDU_R - RDU_L)/\mu \quad (3''')$$

instead of (3'''). The conditional log-likelihood, ignoring indifference, becomes

$$\ln L^{RDU}(\rho, \gamma, \mu; y, \mathbf{X}) = \sum_i l_i^{RDU} = \sum_i [(\ln \Phi(\nabla RDU)) \times \mathbf{I}(y_i=1)) + (\ln (1 - \Phi(\nabla RDU))) \times \mathbf{I}(y_i = -1)] \quad (5'')$$

and requires the estimation of ρ , γ and μ .

For RDEV one replaces (2') with a specification that weights the prizes themselves, rather than the utility of the prizes:

$$RDEV_i = \sum_{k=1,K} [w_k \times m_k] \quad (2'')$$

where m_k is the k^{th} monetary prize. In effect, the RDEV specification is a special case of RDU with the constraint $\rho=1$.

We illustrate the effects of allowing for probability weighting in Panel D of Table 1. When we estimate the RDU model using these data and specification, we find virtually no evidence of probability weighting. The estimate of γ is 0.986 with a 95% confidence interval between 0.971 and 1.002. The hypothesis that $\gamma=1$, that there is no probability weighting, has a p -value of 0.096 using a Wald test. The estimate of the curvature of the utility function, given by ρ , is virtually the same as the estimate of that curvature under EUT in the comparable specification in Panel B ($\rho=0.763$ and $r=0.771$). The effect of allowing for probability weighting is therefore to make no significant change to estimates of the curvature of the utility function – we should be careful here not to *conceptually* associate curvature of the utility function with risk aversion, even if they have essentially the same empirical value in this case.

It is perhaps not surprising, given the precision of the estimate of ρ , that one can easily reject the hypothesis that behavior is consistent with the RDEV model. If one does impose the estimation constraint $\rho=1$, the estimate of γ becomes 1.01, again indistinguishable from EUT.

Of course, these estimates do not support the general claim that probability weighting is irrelevant. These are simply consistent estimates of a structural model given one set of (popular) functional forms and one (large) set of observed responses to diverse lottery choices. Changes in

either might affect estimates significantly.¹¹

3. Loss Aversion and Sign-Dependent Utility

3.1 Original Prospect Theory

Kahneman and Tversky [1979] introduced the notion of sign-dependent preferences, stressing the role of the reference point when evaluating lotteries. In various forms, as we will see, Prospect Theory (PT) has become the most popular alternative to EUT. Original Prospect Theory (OPT) departs from EUT in three major ways: (a) allowance for subjective probability weighting; (b) allowance for a reference point defined over outcomes, and the use of different utility functions for gains or losses; and (c) allowance for loss aversion, the notion that the disutility of losses weighs more heavily than the utility of comparable gains.

The first step is probability weighting, of the form $\omega(p)$ defined in (7), for example. One of the central assumptions of OPT, differentiating it from later variants of PT, is that $w(p) = \omega(p)$, so that the transformed probabilities given by $\omega(p)$ are directly used to evaluate prospective utility:

$$PU_i = \sum_{k=1, K} [\omega_k \times u_k]. \quad (2'')$$

The second step in OPT is to define a reference point so that one can identify outcomes as gains or losses. Let the reference point be given by χ for a given subject in a given round. Consistent with the functional forms widely used in PT, we again use the CRRA functional form

$$u(m) = m^\alpha \quad (1''')$$

when $m \geq \chi$, and

$$u(m) = -\lambda[(-m)^\alpha] \quad (1''')$$

when $m < \chi$, and where λ is the loss aversion parameter. We use the same exponent α for the utility functions defined over gains and losses, even though the original statements of PT keep them theoretically distinct; β is normally used as the counterpart to α when $m < \chi$. Köbberling and

¹¹ For example, dropping the Fechner error specification results in no noticeable change in ρ , but γ drops slight to 0.93, with a 95% confidence interval between 0.88 and 0.97. The p -value on the hypothesis that $\gamma=1$ drops to 0.0015, so there is now statistically significant evidence of probability weighting, even if it is not quantitatively large.

Wakker [2005; §7] point out that the constraint $\alpha=\beta$ is needed to identify the degree of loss aversion if one uses CRRA functional forms *and* does not want to make other strong assumptions (e.g., that utility is measurable only on a ratio scale).¹² Although λ is free in principle to be less than 1 or greater than 1, most PT analysts presume that $\lambda \geq 1$, and we can either impose this as an estimating constraint¹³ if we believe dogmatically in that prior, or we can evaluate it. For the moment, we assume that the reference point is provided by the experimenter-induced frame of the task, and that λ is unconstrained.

The reference point also influences the nature of subjective probability weighting assumed, since different weights are often allowed for gains and losses. Thus we again assume

$$\omega(p) = p^\gamma / [p^\gamma + (1-p)^\gamma]^{1/\gamma} \quad (7)$$

for gains, but estimate

$$\omega(p) = p^\phi / [p^\phi + (1-p)^\phi]^{1/\phi} \quad (7')$$

for losses. It is common in empirical applications to assume $\gamma=\phi$, and we make this assumption as well in our estimation examples for simplicity.

The remainder of the econometric specification is the same as for the EUT and RDU model. The latent index can be defined in the same manner, and the conditional log-likelihood defined comparably. Estimation of the core parameters α , λ , γ , ϕ and μ is required.

The primary logical problem with OPT was that it implied violations of stochastic dominance. Whenever $\gamma \neq 1$ or $\phi \neq 1$, it is possible to find non-degenerate lotteries such that one

¹² The estimates of the coefficient obtained by Tversky and Kahneman [1992] fortuitously happened to be the same for losses and gains, and many applications of PT assume that for convenience. The empirical methods of Tversky and Kahneman [1992] are difficult to defend, however: they report median values of the *estimates* obtained after fitting their model for each subject. The estimation for each subject is attractive if data permits, as magnificently demonstrated by Hey and Orme [1994], but the *median estimate* has nothing to commend it statistically. Within psychology, Birnbaum and Chavez [1997] also estimate at the level of the individual, but then report the median estimate for each parameter over 100 subjects. Their estimation approach is actually maximum likelihood if the parameter h in their objective function (20) is set to 0; in fact, it is set to 0.01 in the reported estimates, which effectively makes these maximum likelihood estimates. Unfortunately their estimation procedure does not seem to generate standard errors.

¹³ Inequality constraints are handled by estimating a parameter that is some non-linear transform of the parameter of interest, but that can vary between $\pm\infty$ to allow gradient-based algorithms free rein. For example, to impose a non-negativity constraint on some deep parameter η one could estimate a parameter $\hat{\eta}$ and define η inside the likelihood evaluator as $\eta = \ln(\hat{\eta})$. Similarly, to constrain the parameter η to some interval (a, b) , for given values of a and b , one could estimate $\hat{\eta}$ and define η inside the likelihood evaluator by the logistic transform $\eta = a + b/(1+\exp(\hat{\eta}))$. Estimates of $\hat{\eta}$ are returned by the maximum likelihood evaluator, and one can infer point estimates and standard errors for η using the “delta method” (Oehlert [1992]).

lottery would stochastically dominate the other, but would be assigned a lower PU. Examples arise quickly when one recognizes that $\gamma(p_1 + p_2) \neq \gamma(p_1) + \gamma(p_2)$ for some p_1 and p_2 . Kahneman and Tversky [1979] dealt with this problem by assuming that evaluation using OPT only occurred after dominated lotteries were eliminated. When not including an editing phase in the structural estimation one may interpret the stochastic error term μ as a reduced form proxy for that editing process.¹⁴

3.2 Cumulative Prospect Theory

The notion of rank-dependent decision weights was incorporated into OPT by Starmer and Sugden [1989], Luce and Fishburn [1991] and Tversky and Kahneman [1992]. Instead of implicitly assuming that $w(p) = \omega(p)$, it allowed $w(p)$ to be defined as in the RDU specification (6a) and (6b). The sign-dependence of subjective probability weighting in OPT, leading to the estimation of different probability weighting functions (7) and (7') for gains and losses, is maintained in Cumulative Prospect Theory (CPT). Thus there is a separate decumulative function used for gains and losses, but otherwise the logic is the same as for RDU.¹⁵

The estimation of a structural CPT model is illustrated with the same data used for EUT and RDU. Here we allow the experimenter-induced frame to define what is a gain and a loss, but we will discuss below how important the assumption of the reference point is to the estimation. Panels E and F show maximum likelihood estimates of the simplified CPT model in which $\gamma = \phi$. In one case (Panel F) we impose the further constraint that $\alpha = \beta$, for the theoretical reasons noted above.

Focusing on the unconstrained estimates in Panel E, we estimate a concave utility function over

¹⁴ In other words, evaluating the PU of two lotteries, without having edited out dominated lotteries, might lead to a dominated lottery having a higher PU. But if subjects always reject dominated lotteries, the choice would appear to be an error to the likelihood function. Apart from searching for better parameters to explain this error, as the maximum likelihood algorithm does as it tries to find parameter estimates that reduce any other prediction error, our specification allows μ to increase. We stress that this argument is not intended to rationalize the use of separable probability weights in OPT, just to explain how a structural model with stochastic errors might account for the effects of stochastic dominance. Wakker [1989] contains a careful account of the notion of transforming probabilities in a “natural way” but without violating stochastic dominance.

¹⁵ One of the little secrets of CPT is that one must always have a probability weight for the residual outcome associated with the reference point, and that the reference outcome receive a utility of 0 for both gains and losses. This ensures that decision weights always add up to 1.

gains ($\alpha < 1$), a convex utility function over losses ($\beta < 1$), evidence of loss seeking ($\lambda < 1$) instead of loss aversion, and mild evidence of probability weighting in the expected direction ($\gamma < 1$). The most striking result here is that loss aversion does not leap out: most PT analysts employ the estimate of 2.25 estimated by Tversky and Kahneman [1992] without question. These illustrative findings should give some reason to pause and question such an assumption.

The constrained estimates in Panel F are similar, but exhibit roughly the same degree of concavity in the gain domain and less convexity in the loss domain ($\alpha = 0.582 = \beta$, compared to $\alpha = 0.535$ and $\beta = 0.930$ from Panel E). The extent of loss seeking is mitigated slightly, but still there is no evidence of loss aversion: the 95% confidence interval for λ is between 0.778 and 1.007. It is noteworthy that the addition of the constraint $\alpha = \beta$ reduced the log-likelihood, and indeed one can formally reject the hypothesis that $\alpha = \beta$ on empirical grounds using the estimates from the unconstrained model in Panel E: the χ^2 statistic has a value of 31.9, so with 1 degree of freedom the p -value on this test is less than 0.0001. On the other hand, there is a significant theoretical trade-off if one maintains this difference between α and β , stressed by Köbberling and Wakker [2005; §7], so this is not the sort of constraint that one should decide on purely empirical grounds.

3.3 Will the True Reference Point Please Stand Up?

It is essential to take a structural perspective when estimating CPT models. Estimates of the loss aversion parameter depend intimately on the assumed reference point, as one would expect since the latter determines what are to be viewed as losses. So if we have assumed the wrong reference point, we will not reliably estimate the degree of loss aversion. However, if we do not get loss aversion leaping out at us when we make a natural assumption about the reference point, should we infer that there is no loss aversion or that there is loss aversion and we just used the wrong reference point? This question points to a key operational weakness of CPT: the need to specify what the reference point is. Loss aversion *may* be present for *some* reference point, but if it is not present for the one we used, and none others are “obviously” better, then should one keep searching

for some reference point that generates loss aversion? Without a convincing argument about the correct reference point, and evidence for loss aversion conditional on that reference point, one simply cannot claim that loss aversion is always present. This specification ambiguity is arguably less severe in the lab, where one can frame tasks to try to induce a loss frame, but is a particularly serious issue in the field.

Similarly, estimates of the nature of probability weighting vary with changes in reference points, loss aversion parameters, and the concavity of the utility function, and *vice versa*. All of this is to be expected from the CPT model, but necessitates joint econometric estimation of these parameters if one is to be able to make consistent statements about behavior.

In many laboratory experiments it is simply assumed that the manner in which the task is framed to the subject defines the reference point that the subject uses. Thus, if one tells the subject that they have an endowment of \$15 and that one lottery outcome is to have \$8 taken from them, then the frame might be appropriately assumed to be \$15 and this outcome coded as a loss of \$8. But if the subject had been told, or expected, that they would earn only \$5 from the experimental task, would this be coded instead as a gain of \$2 (since the \$15 endowment net of the \$8 loss is earnings of \$7, thus \$2 more than expected)? The subjectivity and contextual nature of the reference point has been emphasized throughout by Kahneman and Tversky [1979], even though one often collapses it to the experimenter-induced frame in evaluating laboratory experiments. This imprecision in the reference point is not a criticism of PT, just a challenge to be careful assuming that it is always fixed and deterministic (see Kőszegi and Rabin [2006][2007], Schmidt, Starmer and Sugden [2008] and Harrison and Rutström [2008; §3.2.3]).

A corollary is that it might be a mistake to view loss aversion as a fixed parameter λ that does not vary with the context of the decision, *ceteris paribus* the reference point. This concern is discussed by Novemsky and Kahneman [2005] and Camerer [2005; p.132, 133], and arises most clearly in dynamic decision-making settings with path-dependent earnings.

To gauge the extent of the problem, we re-visit the estimation of a structural CPT model

using our laboratory data, but this time consider the effect of assuming different reference points than the one induced by the task frame. Assume that the reference point is χ , as in (1''') and (1''') above, but instead of setting $\chi = \$0$, allow it to vary between \$0 and \$10 in increments of \$0.10. The results are displayed in Figure 3. The top left panel shows a trace of the log-likelihood value as the reference point is increased, and reaches a maximum at \$4.50. To properly interpret this value, note that these estimates are made at the level of the individual choice in this task, and the subject was to be paid for 3 of those choices. So the reference point for the overall task of 60 choices would be \$13.50 ($=3 \times \4.50). This is roughly consistent with the range of estimates of expected session earnings elicited by Andersen, Harrison and Rutström [2006] for a sample drawn from the same population.¹⁶ It is also significantly higher than the show-up fee for these sessions, which was \$5.

The other interesting part of Figure 3 is that the estimate of loss aversion increases steadily as one increases the assumed reference point. At the maximum likelihood reference point of \$4.50, λ is estimated to be 2.72, with a standard error of 0.42 and a 95% confidence interval between 1.90 and 3.54. We now see that our earlier estimation of a loss aversion coefficient that is less than one (as shown in Panel F or Table 1) is a function of the assumed reference point at \$0. As soon as this reference point is assumed to exceed \$1 we would estimate λ greater than 1, consistent with aversion to losses.

The lesson here is that estimates of loss aversion depend crucially on the priors held by the analyst. Adopting particular parameter values for λ without acknowledging its relationship to these priors is at best a casual practice and can lead to biases in estimation of other key parameters.

4. Mixture Models and Multiple Decision Processes

4.1 Recognizing Multiple Decision Processes

Since different models of choice behavior under uncertainty affect the characterization of

¹⁶ The mean estimate from their sample was \$31, but there were clear nodes at \$15 and \$30. Our experimental sessions typically consist of several tasks, so expected earnings from the lottery task would have been some fraction of these expectations over session earnings. No subject stated an expected earning below \$7.

risk attitudes, it is of some importance that we make some determination about which of these models is to be adopted. One of the enduring contributions of behavioral economics is that we now have a rich set of competing models of behavior in many settings, with EUT and PT as the two front runners for choices under uncertainty. Debates over the validity of these models have often been framed as a horse race, with the winning theory being declared on the basis of some statistical test in which the theory is represented as a latent process explaining the data. In other words, we seem to pick the best theory by “majority rule.” If one theory explains more of the data than another theory, we declare it the better theory and discard the other one. In effect, after the race is over we view the horse that “wins by a nose” as if it was the only horse in the race. The problem with this approach is that it does not recognize the possibility that several behavioral latent processes may coexist in a population. Recognizing that possibility has direct implications for the characterization of risk attitudes in the population.

Ignoring this possibility can lead to erroneous conclusions about the domain of applicability of each theory, and is likely an important reason for why the horse races pick different winners in different domains. For purely statistical reasons, if we have a belief that there are two or more latent population processes generating the observed sample, one can make more appropriate inferences if the data are not forced to fit a specification that assumes one latent population process.

Heterogeneity in responses is well recognized as causing statistical problems in experimental and non-experimental data. Nevertheless, allowing for heterogeneity in responses through standard methods, such as fixed or random effects, is not helpful when we want to identify which people behave according to what theory, and when. Heterogeneity can be partially recognized by collecting information on observable characteristics and controlling for them in the statistical analysis. For example, a given theory might allow some individuals to be more risk averse than others as a matter of personal preference. But this approach only recognizes *heterogeneity within a given theory*. This may be important for valid inferences about the ability of the theory to explain the data, but it does not allow for *heterogeneous theories* to co-exist in the same sample.

The approach to heterogeneity and the possibility of co-existing theories adopted by Harrison and Rutström [2009] is to propose a “wedding” of the theories. They specify and estimate a grand likelihood function that allows each theory to co-exist and have different weights, a so-called mixture model. The data can then identify what support each theory has. The wedding is consummated by the maximum likelihood estimates converging on probabilities that apportion non-trivial weights to each theory.

Their results are striking: EUT and PT share the stage, in the sense that each accounts for roughly 50% of the observed choices. Thus, to the extent that EUT and PT imply different things about how one measures risk aversion, and the role of the utility function as against other constructs, assuming that the data is generated by one or the other model can lead to erroneous conclusions. The fact that the mixture probability is estimated with some precision, and that one can reject the null hypothesis that it is either 0 or 1, also indicates that one cannot claim that the equal weight to these models is due to chance.

Andersen, Harrison, Lau and Rutström [2008a] apply the same notion of mixture models to consider the possibility that discounting behavior in experiments is characterized by a combination of exponential and hyperbolic specifications. They find that the exponential model accounts for roughly 72% of the observed choices, but that one cannot reject the hypothesis that both processes were operating. That is, even if the exponential model “wins” in the sense that 72% is greater than 50%, the correct specification includes both processes.

The main methodological lesson from these exercises is that one should not rush to declare one or other model as a winner in all settings. One would expect that the weight attached to EUT or the exponential model of discounting would vary across task domains, just as it can be shown to vary across observable socio-economics characteristics of individual decision makers.

4.2 Implications for the Interpretation of Process Data

An important tradition in psychology uses data on the processes of decision to discriminate

between models. The earliest traditions no doubt stem from casual introspection, but formal developments include the use of verbal protocols advocated by Ericsson and Simon [1993]. Apart from the assumption that the collection of process data does not affect the process used, inference using process data would seem to require that some *a priori* restriction be placed on the decision-making processes admitted.

For example, Johnson, Schulte-Mecklenbeck and Willemsen [2008] present evidence that subjects in lottery choice settings evaluate tradeoffs between probabilities and prizes as they roll their mouse around a screen to gather up these crumbs of data on a lottery. This may be of some value in suggesting that models that rule out such tradeoffs, such as the priority heuristic of Brandstätter, Gigerenzer and Hertwig [2006], do not explain *all* of the data. But they do not rule out the notion that such heuristics play a role after some initial phase in which subjects determine if the expected value of one lottery vastly exceeds the other, as Brandstätter, Gigerenzer and Hertwig [2006; p.425ff.] allow in an important qualification (see §6 below for more details on the priority heuristic). Nor does it allow one to rule out hypotheses that models such as the priority heuristic might be used by a given subject in mixtures with more traditional models, such as EUT.

Since we stress the interpretation of formal models in terms of latent processes, we would never want to discard data that purports to reflect that process. But the assumptions needed to make those connections are, as yet, heroic and speculative, as some of the wilder claims of the neuroeconomics literature demonstrate all too well.

4.3 Comparing Latent Process Models

Whenever one considers two competing models, readers expect to see some comparative measures of goodness of fit. Common measures include R^2 , pseudo- R^2 , a “hit ratio,” some other scalar appropriate for choice models (e.g., Hosmer and Lemeshow [2000; ch.5]), and formal likelihood-ratio tests of one model against another (e.g., Cox [1961][1962] or Vuong [1989]). From the perspective adopted here, the *interpretation* of these tests suffers from the problem of implicitly

assuming just one data-generating process. In effect, the mixture model provides a built-in comparative measure of goodness of fit – the mixture probability itself. If this probability is close to 0 or 1 by standard tests, one of the models is effectively rejected, in favor of the hypothesis that there is just one data-generating process.

In fact, if one traces back through the literature on non-nested hypothesis tests, these points are “well known.” That literature is generally held to have been started formally by Cox [1961], who proposed a test statistic that generalized the usual likelihood ratio test (LRT). His test compares the difference between the actual LRT of the two models with the expected LRT, suitably normalized by the variance of that difference, under the hypothesis that one of the models is the true data-generating process. The statistic is applied symmetrically to both models, in the sense that each takes a turn at being the true model, and leads to one of four conclusions: one model is the true model, the other model is the true model, neither model is true, or both models are true.¹⁷

However, what is often missed is that Cox [1962; p.407] briefly, but explicitly, proposed a multiplicative mixture model as an “alternative important method of tackling these problems.” He noted that this “procedure has the major advantage of leading to an estimation procedure as well as to a significance test. Usually, however, the calculations will be very complicated.” Given the computational limitations and costs of the day, he did not pursue the mixture model approach further since it would have been inefficient to do so.

The next step in the statistical literature was the development by Atkinson [1970] of the suggestion of Cox. The main problem with this exposition, noted by virtually every commentator in the ensuing discussion, was the interpretation of the mixing parameter. Atkinson [1970; p.324] focused on testing the hypothesis that this parameter equaled $\frac{1}{2}$, “which implies that both models fit the data equally well, or equally badly.” There is a colloquial sense in which this is a correct interpretation, but it can easily lead to confusion if one maintains the hypothesis that there is only

¹⁷ This possible ambiguity is viewed as an undesirable feature of the test by some, but not when the test is viewed as one of an armada of possible model *specification* tests rather than as a model *selection* tests. See Pollak and Wales [1991; p. 227ff.] and Davidson and MacKinnon [1993; p. 384] for clear discussions of these differences.

one true data generating process, as the commentators do. In that case one is indeed confusing model specification tests with model selection tests. If instead the possibility that there are two data generating processes is allowed, then natural interpretations of tests of this kind arise.¹⁸

Computational constraints again restricted Atkinson [1970] to deriving results for tractable special cases.¹⁹

This idea was more completely developed by Quandt [1974] in the additive mixture form we use. He did, however, add a seemingly strange comment that “The resulting pdf is formally identical with the pdf of a random variable produced by a mixture of two distributions. It is stressed that this is a formal similarity only.” (p.93/4) His point again derives from the tacit assumption that there is only one data generating process rather than two (or more). From the former perspective, he proposes viewing corner values of the mixture probability as evidence that one or other model is the true model, but to view interior values as evidence that some unknown model is actually used and that a mixture of the two proposed models just happens to provide a better approximation to that unknown, true model. But if we adopt the perspective that there are two possible data generating processes, the use and interpretation of the mixing probability estimate is direct.

Perhaps the most popular modern variant of the generalized LRT approach of Cox [1961][1962] is due to Vuong [1989]. He proposes the null hypothesis that both models are the true models, and then allows two one-sided alternative hypotheses.²⁰ The statistic he derives takes *observation-specific* ratios of the likelihoods under each model, so that in our case the ratio for observation i is the likelihood of observation i under EUT divided by the likelihood of observation i

¹⁸ Of course, as noted earlier, there are several possible interpretations in terms of mixtures occurring at the level of the observation (lottery choice) or the unit of observation (the subject or task). Quandt [1974] and Pesaran [1981] discuss problems with the multiplicative mixture specification from the perspective of the data being generated by a single process.

¹⁹ These constraints were even binding on methodology as recently as Pollak and Wales [1991]. They note (p.228) that “If we could estimate the composite {the mixture specification proposed by Atkinson [1970] and Quandt [1974]}, then we could use the standard likelihood ratio test procedure to compare the two hypotheses with the composite and there would be no reason to focus on choosing between the two hypotheses without the option of rejecting them both in favor of the composite. Thus, the model selection problem arises only when one cannot estimate the composite.” They later discuss the estimation problems in their extended example, primarily deriving from the highly non-linear functional form (p.232). As a result, they devise an ingenious method for ranking the alternative models under the *maintained assumption* that one cannot estimate the composite (p.230).

²⁰ Some have criticized the Vuong test because the null hypothesis is often logically impossible, but it can also be interpreted as the hypothesis that one cannot say which model is correct.

under PT. It then calculates the log of these ratios, and tests whether the expected value of these log-ratios over the sample is zero. Under reasonably general conditions a normalized version of this statistic is distributed according to the standard normal, allowing test criteria to be developed.²¹ Thus the resulting statistic typically provides evidence in favor of one of the models which may or may not be statistically significant.

Applying the Vuong test to the EUT and PT models estimated independently by Harrison and Rutström [2009; Table 1], we would conclude that there is overwhelming evidence in favor of the PT model.²² However, when we use the Vuong test of the PT-only model against the mixture model, the test statistic favors the mixture model; the test statistic is -0.56 , with a p -value of 0.71 that the PT-only model is *not* the better model. The inferences that one draws from these test statistics therefore depend critically on the perspective adopted with respect to the data generating process. If we look for a single data generating process in our case, then PT dominates EUT. But if one allows the data to be generated by either model, the evidence is mixed – if one excuses the pun, and correctly interprets that as saying that both models receive roughly the same support. *Thus one would be led to the wrong qualitative conclusion if the non-nested hypothesis tests had been mechanically applied.*

5. Dual Criterion Models from Psychology

The prevailing approach of economists to this problem is to assume a single criterion, whether it reflects standard EUT, RDU, or PT. In each case the risky prospect is reduced to some scalar, representing the preferences, framing and budget constraints of the decision-maker, and then that scalar is used to rank alternatives.²³

²¹ Clarke [2003] proposes a non-parametric sign test be applied to the sample of ratios. Clarke [2007] demonstrates that when the distribution of the log of the likelihood ratios is normally distributed that the Vuong test is better in terms of asymptotic efficiency. But if this distribution exhibits sharp peaks, in the sense that it is mesokurtic, then the non-parametric version is better. The likelihood ratios we are dealing with have the latter shape.

²² The test statistic has a value of -10.33 . There are often additional corrections for degrees of freedom, using one or other “information criteria” to penalize models with more parameters (in our case, the PT model). We do not accept the underlying premiss of these corrections, that smaller models are better, and do not make these corrections. The results reported below would be the same if we did.

²³ In economics the only exceptions are lexicographic models, although one might view the criterion at each stage as being contemplated simultaneously. For example, Rubinstein [1988] and Leland [1994] consider the use of similarity relations in conjunction with “some other criterion” if the similarity relation does not recommend a choice. In

Many other disciplines assume the use of decision-making models with multiple criteria.²⁴ In some cases these models can be reduced to a single criterion framework, and represent a recognition that there may be many attributes or arguments of that criterion. Other times decision rules provide different metrics for evaluating what to do, and all aspects of a decision may not reduce to a single-criterion.

An alternative decision rule is provided by the SP/A model of Lopes [1995]. This model departs from EUT, RDU and PT in one major respect: it is a dual criterion model. Each of the single criterion models, even if they have a number of components to their evaluation stage, boil down to a scalar index for each lottery such as (2), (2') and (2''). The SP/A model instead explicitly posits two distinct but simultaneous ways in which the same subject might evaluate a given lottery. One is the SP part, for a process that weights the “security” and “potential” of the lottery in ways that are similar to RDEV. The other is the A part, which focuses on the “aspirations” of the decision-maker. In many settings these two parts appear to be in conflict, which means that one must be precise as to how that conflict is resolved. We discuss each part, and then how the two parts may be jointly estimated.

5.1 The “SP” Criterion

Although motivated differently, the SP criterion is formally identical to the RDEV criterion reviewed earlier. The decision weights in SP/A theory derive from a process by which the decision-maker balances the security and potential of a lottery. On average, the evidence collected from experiments, such as those described in Lopes [1984], seems to suggest that an inverted-S shape familiar from PT

fact, Rubinstein [1988] and Leland [1994] reverse the sequential order in which the two criteria are applied, indicating some sense of uncertainty about the strict sequencing of the application of criteria. Similarly, the original prospect theory of Kahneman and Tversky [1979] considered an “editing stage” to be followed by an “evaluation stage,” although the former appears to have been edited out of later variants of prospect theory.

²⁴ Quite apart from the model from psychology evaluated here, there is a large literature in psychology referenced by Starmer [2000] and Brandstätter, Gigerenzer and Hertwig [2006]. Andersen, Harrison, Lau and Rutström [2006] review the various models in other disciplines.

... represents the weighting pattern of the average decision maker. The function is security-minded for low outcomes (i.e., proportionally more attention is devoted to worse outcomes than to moderate outcomes) but there is some overweighting (extra attention) given to the very best outcomes. A person displaying the cautiously hopeful pattern would be basically security-minded but would consider potential when security differences were small. (Lopes [1995; p.186])

The upshot is that the probability weighting function

$$\omega(p) = p^\gamma / [p^\gamma + (1-p)^\gamma]^{1/\gamma} \quad (7)$$

from RDU would be employed by the *average* subject, with the expectation that $\gamma < 1$.²⁵ However, there is no presumption that any individual subject follow this pattern. Most presentations of the SP/A model assume that subjects use a linear utility function, but this is a convenience more than anything else. Lopes and Oden [1999; p.290] argue that

Most theorists assume that [utility] is linear without asking whether the monetary range under consideration is wide enough for nonlinearity to be manifest in the data. We believe that [utility] probably does have mild concavity that might be manifest in some cases (as, for example, when someone is considering the huge payouts in state lotteries). But for narrower ranges, we prefer to ignore concavity and let the decumulative weighting function carry the theoretical load.

So the SP part of the SP/A model collapses to be the same as RDU, although the interpretation of the probability weighting function and decision weights is quite different. The restriction to the RDEV model suggested by Lopes and Oden [1999] can then be tested empirically, depending on the domain of the tasks used for estimation. Thus we obtain the likelihood of the observed choices conditional on the SP criterion being used to explain them; the same latent index (3'') is constructed, and the likelihood is then (5'') as with RDU. The typical element of that log-likelihood for observation *i* can be denoted l_i^{SP} .

5.2 The "A" Criterion

The aspiration part of the SP/A model collapses the indicator of the value of each lottery down to an expression showing the extent to which it satisfies the aspiration level of the contestant.

²⁵ Lopes and Oden [1999; equation (10), p.290] propose an alternative function which would provide a close approximation to (7). Their function is a weighted average of a convex and concave function, which allows them to interpret the average inverted S-pattern in terms of a weighted mixture of security-minded subjects and potential-minded subjects.

This criterion is sign-dependent in the sense that it defines a threshold for each lottery: if the lottery exceeds that threshold, the subject is more likely to choose it. If there are up to K prizes, then this indicator is given by

$$A_i = \sum_{k=1, K} [\eta_k \times p_k] \quad (8)$$

where η_k is a number that reflects the degree to which prize m_k satisfies the aspiration level, so η_k in effect substitutes for the utility function in EUT or RDEU. The parameter p_k is again the probability of the lottery paying m_k . Although Oden and Lopes [1997] advance an alternative interpretation using fuzzy set theory, so that η_k measures the degree of membership in the set of prizes that are aspired to, we can view this as simply an indicator function, like a utility function. It could be viewed as a crisp, binary threshold for the *individual* subject, which is consistent with it being modelled as a smooth, probabilistic threshold for a *sample* of subjects, as here. Thus, like probabilities, the indicator function η_k is normalized to lie in the unit interval.

This concept of aspiration levels is close to the notion of a threshold income level debated by Camerer, Babcock, Loewenstein and Thaler [1997] and Farber [2005]. The concept is also reminiscent of the “safety first” principle proposed by Roy [1952][1956] and the “confidence limit criterion” of Baumol [1963], although in each case these are presented as extensions of an expected utility criterion rather than as alternatives.

The implication of (8) is that one has to estimate some function mapping prizes into probabilities, to reflect the aspirations of the decision-maker. We use an extremely flexible function for this, the cumulative non-central Beta distribution defined by Johnson, Kotz and Balakrishnan [1995]. This function has three parameters, χ , ξ and ψ . We employ a flexible form simply because we have no *a priori* restrictions on the shape of this function, other than those of a cumulative density function, and in the absence of theoretical guidance prefer to let the data determine these values.²⁶ We want to allow it to be a step function, in case the average decision-maker has some crisp focal point such as £25,000, but the function should then determine the value of the focal point

²⁶ It also helps that this function can be evaluated as an intrinsic function in advanced statistical packages such as *Stata*.

(hence the need for a non-central distribution, given by the parameter ψ). But we also want to allow it to have an inverted S-shape in the same sense that a logistic curve might, or to be convex or concave over the entire domain (hence the two parameters χ and ξ).

Once we have values for η_k it is a simple matter to evaluate A_i using (8). We then construct the likelihood of the data assuming that this criterion was used to explain the observed choices. The likelihood, conditional on the A criterion being the one used by the decision maker, and our functional form for η_k , depends on the estimates χ, ξ, ψ and μ given the above specification and the observed choices. The conditional log-likelihood is

$$\ln L^A(\chi, \xi, \psi, \mu; y) = \sum_i l_i^A = \sum_i [(\ln G(\nabla A)) \times \mathbf{I}(y_i = 1) + (\ln (1-G(\nabla A))) \times \mathbf{I}(y_i = -1)] \quad (9)$$

in the usual manner.

5.3 Combining the Two Criterion

There is a deliberate ambiguity in the manner in which the SP and A criteria are to be combined to predict a specific choice. One reason is a desire to be able to explain evidence of intransitivities, which figures prominently in the psychological literature on choice (e.g., Tversky [1969]). Another reason is the desire to allow context to drive the manner in which the two criteria are combined, to reconcile the model of the choice process with evidence from verbal protocols of decision makers in different contexts. Lopes [1995; p.214] notes the SP/A model can be viewed as a function F of the two criteria, SP and A, and that it

... combines two inputs that are logically and mathematically distinct, much as Allais [1979] proposed long ago. Because SP and A provide conceptually independent assessments of a gamble's attractiveness, one possibility is that F is a weighted average in which the relative weights assigned to SP and A reflect their relative importance in the current decision environment. Another possibility is that F is multiplicative. In either version, however, F would yield a unitary value for each gamble, in which case SP/A would be unable to predict the sorts of intransitivities demonstrated by Tversky [1969] and others.

These proposals involve creating a unitary index of the relative attractiveness of one lottery over another, of the form

$$\nabla_{SP/A} = [\theta \times \nabla_{SP}] + [(1-\theta) \times \nabla_A] \quad (10)$$

for example, where θ is some weighting constant that might be assumed or estimated.²⁷ This scalar measure might then be converted into a cumulative probability $G(\nabla\text{SP}/\text{A}) = \Phi(\nabla\text{SP}/\text{A})$ and a likelihood function defined.

A more natural formulation is provided by thinking of the SP/A model as a mixture of two distinct latent, data-generating processes. If we let π^{SP} denote the probability that the SP process is correct, and $\pi^{\text{A}} = (1 - \pi^{\text{SP}})$ denote the probability that the A process is correct, the grand likelihood of the SP/A process as a whole can be written as the probability weighted average of the conditional *likelihoods*. Thus the likelihood for the overall SP/A model is defined as

$$\ln L(\rho, \gamma, \chi, \xi, \psi, \mu, \pi^{\text{SP}}; y, \mathbf{X}) = \sum_i \ln [(\pi^{\text{SP}} \times l_i^{\text{SP}}) + (\pi^{\text{A}} \times l_i^{\text{A}})]. \quad (11)$$

This log-likelihood can be maximized to find estimates of the parameters of each latent process, as well as the mixing probability π^{SP} . The literal interpretation of the mixing probabilities is at the level of the observation, which in this instance is the choice between the relatively safer and riskier of two options, such as lotteries or gambles. In the case of the SP/A model this is a natural interpretation, reflecting two latent psychological processes for a given individual and task.

This approach assumes that any one observation can be generated by both criteria, although it admits of extremes in which one or other criterion wholly generates the observation. One could alternatively define a grand likelihood in which observations or subjects are classified as following one criterion or the other on the basis of the latent probabilities π^{SP} and π^{A} . El-Gamal and Grether [1995] illustrate this approach in the context of identifying behavioral strategies in Bayesian updating experiments. However, in the case of the SP/A model, it is natural to view the tension between the criteria as reflecting the decisions of a given individual for a given task. Thus we do not believe it would be consistent with the SP/A model to categorize *choices* as wholly driven either by SP or A.

These priors also imply that we prefer not to use mixture specifications in which *subjects* are categorized as completely SP or A. It is possible to rewrite the grand likelihood (11) such that $\pi_i^{\text{SP}} = 1$ and $\pi_i^{\text{A}} = 0$ if $l_i^{\text{SP}} > l_i^{\text{A}}$, and $\pi_i^{\text{SP}} = 0$ and $\pi_i^{\text{A}} = 1$ if $l_i^{\text{SP}} < l_i^{\text{A}}$, where the subscript i now refers to

²⁷ Lopes and Oden [1999; equation 16, p.302] offer a multiplicative form which has the same implication of creating one unitary index of the relative attractiveness of one lottery over another.

the individual *subject*. The general problem with this specification is that it assumes that there is no effect on the probability of SP and A from task domain. We do not want to impose that assumption, even for a relatively homogenous task design such as ours.

5.4 Applications

Andersen, Harrison, Lau and Rutström [2006] apply this specification of the SP/A model to observed behavior in the natural game-show experiments *Deal Or No Deal (DOND)*. That game show involves dramatically skewed distributions of prizes, and large stakes. It also involves fascinating issues of modeling dynamic sequences of choices, reviewed in detail by Andersen, Harrison, Lau and Rutström [2008b]. Using data from 461 contestants on the British version of the show, with prizes ranging from 1p up to £250,000, they find that the weight given to the SP criterion is 0.35, with a 95% confidence interval between 0.30 and 0.40. Thus subjects give greater weight to the A criterion, but there is clear evidence that both are employed: the p -value on the hypothesis that the weight on the SP criterion is 0 is less than 0.0001. They also report evidence of concave utility, and modest levels of probability weighting in the predicted direction.

We implemented laboratory versions of the *DOND* game, to complement the natural experimental data from the game shows. The instructions are reproduced in Appendix C.²⁸ They were provided by hand and read out to subjects to ensure that every subject took some time to digest them. As far as possible, they rely on screen shots of the software interface that the subjects were to use to enter their choices. The opening page for the common practice session in the lab, shown in Figure 4, provides the subject with basic information about the task before them, such as how many boxes there were in all and how many boxes needed to be opened in any round.²⁹ In the

²⁸ Appendix C is available in the working paper version available from the authors.

²⁹ The screen shots provided in the instructions and computer interface were much larger, and easier to read. Post, van den Assem, Baltussen and Thaler [2008] also conducted laboratory experiments patterned on *DOND*. They used instructions which were literally taken from the instructions given to participants in the Dutch *DOND* game show, with some introductory text from the experimenters explaining the exchange rate between the experimental game show earnings and take home payoffs. Their approach has the advantage of using the wording of instructions used in the field. Our objective was to implement a laboratory experiment based on the *DOND* task, and clearly referencing it as a natural counterpart to the lab experiment. But we wanted to use instructions which we had complete control over. We wanted subjects to know exactly what bank offer function was going to be used. In our view the two types of *DOND* laboratory

default setup the subject was given the same frame as in the Australian and US versions of this game show: this has more prizes (26 instead of 22) and more rounds (9 instead of 6) than the UK version.

After clicking on the “Begin” box, the lab subject was given the main interface, shown in Figure 5. This provided the basic information for the *DOND* task. The presentation of prizes was patterned after the displays used on the actual game shows. The prizes are shown in the same nominal denomination as the Australian daytime game show, and the subject told that an exchange rate of 1000:1 would be used to convert earnings in the *DOND* task into cash payments at the end of the session. Thus the top cash prize the subject could earn was \$200 in this version.

The subject was asked to click on a box to select “his box,” and then round 1 began. In the instructions we illustrated a subject picking box #26, and then 6 boxes, so that at the end of round 1 he was presented with a deal from the banker, shown in Figure 6. The prizes that had been opened in round 1 were “shaded” on the display, just as they are in the game show display. The subject is then asked to accept \$4,000 or continue. When the game ends the *DOND* task earnings are converted to cash using the exchange rate, and the experimenter prompted to come over and record those earnings. Each subject played at their own pace after the instructions were read aloud.

One important feature of the experimental instructions was to explain how bank offers would be made. The instructions explained the concept of the expected value of unopened prizes, using several worked numerical examples in simple cases. Then subjects were told that the bank offer would be a fraction of that expected value, with the fractions increasing over the rounds as displayed in Figure 7. This display was generated from Australian game show data available at the time. We literally used the parameters defining the function shown in Figure 7 when calculating offers in the experiment, and then rounding to the nearest dollar.³⁰

experiments complement each other, in the same sense in which lab experiments, field experiments and natural experiments are complementary (see Harrison and List [2004]).

³⁰We also implemented a version that reflected the specific format in the UK game show. A subject that participated in the UK treatment would have been given the instructions for the baseline summarized above. They would then play that game out hypothetically as a practice, to become familiar with the task. Then they were presented with instructions for the real task, where the screens referred to are shown in Appendix C2. In the baseline experiments the real task was the same as the practice task, and subjects simply told that. Thus, *every* subject went through the same hypothetical baseline trainer, and some subjects then played it again for real earnings while other subjects played the UK version instead for real earnings.

The subjects for our laboratory experiments were recruited from the general student population of the University of Central Florida in 2006.³¹ We have information on 870 choices made by 125 subjects.³²

Table 2 and Figure 8 display the maximum likelihood estimates obtained using the SP/A model and these laboratory responses. We find some evidence of probability weighting in the lab environment, and we cannot reject the hypothesis that there is some concavity over outcomes in the utility function component of the SP criterion.

However, lab behavior differs in a more fundamental manner than the field behavior: aspiration levels dominate utility evaluations in the SP/A model. We estimate the weight on the SP component, π^{SP} , to be only 0.073, very close to zero. The aspiration level rapidly rises between \$25 and \$100, as shown in Figure 8.³³

6. The Priority Heuristic

One of the valuable contributions of psychology is the focus on the process of decision-making. Economists have tended to focus on the characterization of properties of equilibria, and neglected the connection to explicit or implicit processes that might bring these about. Of course, this was not always so, as the correspondence principle of Samuelson [1947] dramatically illustrated. But it has become a common methodological difference in practice.³⁴ Brandstätter, Gigerenzer and Hertwig [2006] illustrate the extreme alternative, a process model that is amazingly simple and that apparently explains a lot of data. Their “priority heuristic” is therefore a useful case study in the

³¹ Virtually all subjects indicated that they had seen the US version of the game show, which was a major ratings hit on network television in 5 episodes screened daily at prime-time just prior to Christmas in 2005. Our experiments were conducted about a month after the return of the show in the US, following the 2006 Olympic Games.

³² Of this lab sample, 89 participated in the baseline experiments using the Australian and US formats, and 36 participated in the UK version.

³³ Andersen, Harrison and Rutström [2006] provide evidence that subjects drawn from this population come to a laboratory session with some positive expected earnings, quite apart from the show-up fee. Their estimates are generally not as high as \$50 or \$100, but those expectations were elicited before the subjects knew anything about the prizes in the task they were to participate in. Our estimates of the aspiration levels in *DOND* are based on behavior that is fully informed about those prizes.

³⁴ Some would seek to elevate this practice to define what economics is: see Gul and Pesendorfer [2007]. This is simply historically inaccurate and unproductive, quite apart from the debate over the usefulness of “neuroeconomics” that prompted it.

statistical issues considered here, and the role of a maximum likelihood estimation framework applied to a structural model.

The priority heuristic proposes that subjects evaluate binary choices using a sequence of rules, applied lexicographically. For the case of two non-negative outcomes, the heuristic is:

- If one lottery has a minimum gain that is larger than the minimum gain of the other lottery by ω percent or more of the maximum possible gain, pick it.
- Otherwise, if one lottery has a probability of the minimum gain that is at least ω percent better than the other, pick it.
- Otherwise, pick the lottery with the maximum gain.

The parameters ω and ω are each set to 10, based on arguments (p. 412ff.) about “cultural prominence.” The heuristic has a simple extension to consider the probability of the maximum gain when there are more than two outcomes per lottery.

The key feature of this heuristic is that it completely eschews the notion of trading off the utility of prizes and their probabilities.³⁵ This is a bold departure from the traditions embodied in EUT, RDU, PT, and even SP/A theory. What is striking, then, is that it appears to blow *every* other theory out of the water when applied to *every* conceivable decision problem. It explains the Allais Paradox, it explains the Reflection Effect, it explains the Certainty Effect, it explains the Fourfold Pattern, it explains Intransitivities, and it even predicts choices in “diverse sets of choice problems” better than a very long list of alternatives.³⁶

However, there are three problems with the evidence for the priority heuristic.³⁷

³⁵ Of course, there are many such heuristics from psychology and the judgement and decision-making literature, noted explicitly by Brandstätter, Gigerenzer and Hertwig [2006; Table 3, p.417]. These are reviewed in detail in Thorngate [1980] and Payne, Bettman and Johnson [1993].

³⁶ It is notable that the list of opponents arrayed in the dramatic Figures 1 through 5 of Brandstätter, Gigerenzer and Hertwig [2006] do not include EUT with some simple CRRA specification (1) and modest amounts of risk aversion, or even simple EV maximization.

³⁷ Birnbaum [2008] also argues that the data used by Brandstätter, Gigerenzer and Hertwig [2006] was selective. Unfortunately, most of the data he would like to have included is hypothetical (e.g., Birnbaum and Navarrete [1998]) or effectively hypothetical. In some cases data is collected from hundreds of subjects recruited to participate in internet experiments, who were informed that 3 of them (3 out of how many?) would get to play one of their choices for real (e.g., Birnbaum [2004a][2006]). One might argue that much of the data originally used to test the priority heuristic was hypothetical, but one cannot mitigate such problems by just having more such data.

First, one must be extraordinarily careful of claims about “well known stylized facts” about choice, since the behavioral economics literature has become somewhat untethered from the facts in this regard. Simply consider ground zero, the Allais Paradox. It is now well documented that experimental subjects simply do not fall prey to the Allais Paradox like decision-making lemmings when one presents the task for real payments and drops the word “millions” after the prize amount: see Conlisk [1989], Harrison [1994], Burke, Carter, Gominiak and Ohl [1996] and Fan [2002].³⁸ Subjects appear to crank out the EV when given real tasks to perform, and the vast majority behave consistently with EUT as a result.³⁹ This is not to claim that all anomalies or stylized facts are untrue, but there is a casual tendency in the behavioral economics literature to repeatedly assume stylized facts that are simply incorrect. Thus, to return to the Allais Paradox, if the priority heuristic predicts a violation, and in fact the data says otherwise for *motivated* subjects, doesn’t this count directly as evidence *against* the priority heuristic?

The second problem with the evaluation of the performance of the priority heuristic against alternative models is that the *parameters* of those models, when the model relies on parameters, are taken from studies of different subjects and choice tasks. It is as if the CRRA of an EUT model from an Iowa farmer making fertilizer choices had been applied to the portfolio choices of a Manhattan investment banker. The naïve idea is that there is one, true set of parameters that define the model, and that is the model for all time and all domains.⁴⁰ This flies in the face of the default assumption by economists, and not a few psychologists (e.g., Birnbaum [2008]), that individuals might have different preferences over risk. It is notable that many applied researchers disregard that presumption and build tests of theories that assume homogenous preferences, but at least they are

³⁸ It may be well documented, but it is apparently not well known. Birnbaum [2004a] provides a comprehensive review of his own experimental studies of the Allais common consequence paradoxes, does not mention any of the studies referenced here, and then claims as a general matter that using real, credible payments does not affect behavior (p.105).

³⁹ Another concern with many of these stylized examples is that they are conducted on a between-subjects basis, and rely on comparable choices in two pairs of lotteries. Thus one must account for the presumed heterogeneity in risk attitudes when evaluating the statistical power of claims that EUT is rejected. Loomes and Sugden [1998] and Harrison, Johnson, McInnes and Rutström [2007] pay attention to this issue in different ways in their designs.

⁴⁰ There is a folklore joke about how psychologists treat their models the way economists treat their toothbrush: everyone has their own. In this case it seems as though an old, discarded toothbrush is getting passed around to brush data set after data set.

well aware that this is simply an auxiliary assumption made for tractability (e.g., Camerer and Ho [1994; p.186]). In any event, in those instances the researcher at least estimates parameters afresh in some maximum likelihood sense for the sample of interest.

It is a different matter to estimate parameters for a model from responses from a random sample from a given population, and then see if those parameters predict data from another random sample from the *same population*. Although this tends not to be commonly done in economics, it is different than assuming that parameters are universal constants. For example, Birnbaum and Navarrete [1998; p.50] clearly seek to test model predictions “in the manner predicted in advance of the experiment” using parameters from comparable samples. One must take care that the stimuli and recruitment procedures match, of course, so that one is comparing apples to apples.⁴¹ We stress that this issue is not peculiar to psychologists: behavioral economists have an embarrassing tendency to just assume certain critical parameters casually, relying inordinately on the illustrative estimates of Tversky and Kahneman [1992] as noted earlier. For one celebrated example, consider Benartzi and Thaler [1995], who use laboratory-generated estimates from college students to calibrate a model of the behavior of U.S. bond and stock investors. Such exercises are fine as “finger mathematics” exemplars, but are no substitute for estimation on the comparable samples.⁴² In general, economists tend to focus on in-sample comparisons of estimates from different models, although some have considered the formal estimation issues that arise when one seeks to undertake out-of-sample comparisons (Wilcox [2008][2009]). An example would be comparing behavior in one task context to behavior in another task context, albeit a context that is comparable.

The third problem with the priority heuristic is the fundamental one from the present perspective of thinking about models using a maximum likelihood approach: it predicts with

⁴¹ Some degree of robustness in parameters with respect to the stimuli is desirable, but the functional forms used in estimation are often quite restrictive and may not lend themselves to evaluation by universal validity criteria. Researchers need to carefully consider what constitutes appropriate domains for particular functions.

⁴² This example also illustrates the danger of using estimates from one structural model and applying them casually to a different structural model. In this case the prospect theory parameters were held fixed and the best-fitting “evaluation horizon” determined from data. But when one estimates these parameters from responses in controlled experiments in which the evaluation horizon is varied as a treatment, they are not the same (Harrison and Rutström [2008; Appendix E3]).

probability one or zero. So, surely, aren't there *some* interesting settings in which the heuristic must be completely wrong most or all the time? Indeed there are. Consider the comparison of lottery A in which the subject gets \$1.60 with probability p and \$2.00 with probability $1-p$, and lottery B in which the subject gets \$0.10 with probability p and \$3.85 with probability $1-p$. The priority heuristic picks A *every time*, no matter how low p is.⁴³ The minimum gain is \$1.60 for A and \$0.10 for B, and 10% of \$1.60 is \$0.16, greater than \$0.10.

At this point experimental economists are jumping up and down, waving their hands and pointing to the data from a massive range of experiments initiated by Holt and Laury [2002] with exactly these parameters. Their baseline experimental task presented subjects with an ordered list of 10 such choices, with p ranging from 0.1 to 1 in increments of 0.1. Refer to these prizes as their 1x prizes, where the number indicates a scale factor applied to all prizes. Identical tasks are reported by Holt and Laury [2002][2005] with 20x, 50x and 90x prizes, and by Harrison, Johnson, McInnes and Rutström [2005] with 10x prizes. Although we will want to do much, much better than just look at average choices, it is apparent from these data that the priority heuristic must be in trouble as a general model. Holt and Laury [2005; Table 1, p. 903] report that the average number of choices of lottery A is 5.2, 5.3, 6.1 and 5.7 over hundreds of subjects facing the 1x task, 6.0 over 178 subjects facing the 10x task, and 6.7 over 216 subjects facing the 20x task, in all cases for real payments and with no order effects. The predicted outcome for an EUT model assuming risk neutrality is for 4 choices of lottery A, and a modest extension of EUT to allow small levels of risk aversion would explain 5 or 6 safe choices quite well. In fact, using the CRRA utility function (1), any RRA between 0.15 and 0.41 would predict 5 choices, and any RRA between 0.41 and 0.68 would predict 6 choices (Holt and Laury [2002; Table 3, p.1649]).

But using the metric of evaluation of Brandstätter, Gigerenzer and Hertwig [2006] the

⁴³ In fact, there is a threshold ω of the ratio of the expected values of the lotteries, *above* which the priority heuristic is assumed not to apply, and where probabilities and prizes are traded off in the usual EUT manner assuming risk neutrality (p. 425ff). The parameter ω is set to 2, but is apparently not applied in the main tests of the predictive power of alternative theories in Brandstätter, Gigerenzer and Hertwig [2006; p.416ff]. With this modification, the priority heuristic predicts that A might also be selected because of EUT heuristics for $p \leq 0.1972$.

priority heuristic would predict behavior here perfectly as well! This is because they count a success for a theory based on whether it predicts the *majority* choice correctly.⁴⁴ In the 10 choices of the Holt and Laury [2002] task, imagine that subjects picked A on average 5.000000001 times. An EUT model, in which the CRRA was set to around 0.25, would predict that the average subject picks lottery A 5 times and then switches to B for the other 5 choices, hence predicting almost perfectly in each of the 10 choices. But the priority heuristic gets almost 4 out of 10 wrong *every time*, and yet is viewed as a 100% successful theory by this metric.

This example shows exactly why it is a mistake to casually use the “hit rate” as a metric of evaluation in such settings.⁴⁵ The likelihood approach instead asks the model to state the probability of observing the actual choice, conditional on some trial values of the parameters of the theory. Maximum likelihood then just finds those parameters that generate the highest probability of observing the data. For binary choice tasks, and independent observations, we know that the likelihood of the sample is just the product of the likelihood of each choice conditional on the model and the parameters assumed, and that the likelihood of each choice is just the probability of that choice. So if we have any observation that has zero probability, and the priority heuristic has many, the log-likelihood for that observation zooms off to minus infinity. Even if we set the likelihood to some minuscule amount, so we do not have to evaluate the logarithm of zero, the overall likelihood of the priority heuristic is *a priori* abysmal without even firing up any statistical package.

Of course, this is true for any theory that predicts deterministically, including EUT. But this is why one needs some formal statement about how the deterministic prediction of the theory

⁴⁴ To see this follow carefully the explanation in Brandstätter, Gigerenzer and Hertwig [2006; p.418] of how the vertical axis on their Figure 1 is created. There are 14 choice tasks being evaluated here. The priority heuristic predicted the *majority* choice in each of the 14 tasks, so it is given a predictive score of 100%. The “equiprobable” heuristic predicted 10 out of 14 of the *majority* choices, so it is given a predictive score of 71.4% = $(10 \div 14) \times 100$. The predictive accuracy measure is not calculated at the level of the individual choice, but instead using a summary statistic of those choices. Rieger and Wang [2008; Figure 3] make essentially the same point, but do not suggest the preferred evaluation in terms of likelihoods.

⁴⁵ There are some non-casual, semi-parametric estimation procedures for binary choice models that use the hit rate, such as the “maximum score” estimator of Manski [1975]. The literature on this estimator is reviewed by Cameron and Trivedi [2005; §14.7.2, p.483ff.].

translates into a probability of observing one choice or the other, and then perhaps also some formal statement about the role that structural errors might play. Examples include equation (4) from §1.1, and the whole of §1.2, respectively.

How would one modify the priority heuristic to make it worth testing against any real data at an individual level? Perhaps one could count how many of the criteria are pointing towards one lottery, and use that as an indicator of strength of preference. But this path seems *ad hoc*, would need weights on the criterion to avoid discontinuities in any likelihood maximization process using gradient methods, and is contrary to the *raison d'être* of the model.

The priority heuristic is one part of a larger research program into the ecological rationality of “fast and frugal heuristics.” This research program is important because it provides a sharply contrasting view to the prevailing dogma among behavioral economists, that the use of heuristics must always be associated with sub-optimal decisions. In fact, when heuristics have evolved or been consciously tailored to a task domain, they can perform extraordinarily well. When they are applied to the wrong task domain, catastrophic errors can result. These important perspectives from psychology (e.g., Brunswick [1955]) long predated recognition by economists of the role of ecological rationality (cf. Smith [2003]) and artefactual task domains (cf. Harrison and List [2004]). On the other hand, one should not make the case for the descriptive validity of heuristics on flimsy statistical grounds, even if the general approach is an important one in principle.

7. Conclusion

Some psychologists criticize mainstream models because there is alleged evidence that one or more of the implied processes is invalid.⁴⁶ “People do not multiply utilities and probabilities and add them up in their head, since they cannot even do this for artefactual multiplication and addition tasks without a calculator.” The economist retorts that “I only assume that people behave *as if* they do that

⁴⁶ Most claims are strikingly casual. Excellent exceptions are Gigerenzer and Kurz [2001] and Knutson and Peterson [2005].

process.” There is truth in both arguments, and the two views are actually complementary if one separates out intellectual jingoism from substance.⁴⁷ The formal statistical framework proposed here, maximum likelihood of structural models, serves as one tool to actually realize the potential gains from trade between economics and psychology. Much remains to be done.

⁴⁷ The extremes of jingoism arise in the marketing of “neuroeconomics,” as stressed by Harrison [2008]. The more interesting question is the value of having detailed process information in addition to observed choice data, as illustrated by the evaluation of the implied process of the priority heuristic by Johnson, Schulte-Mecklenbeck and Willemsen [2008].

Figure 1: Normal and Logistic Cumulative Density Functions

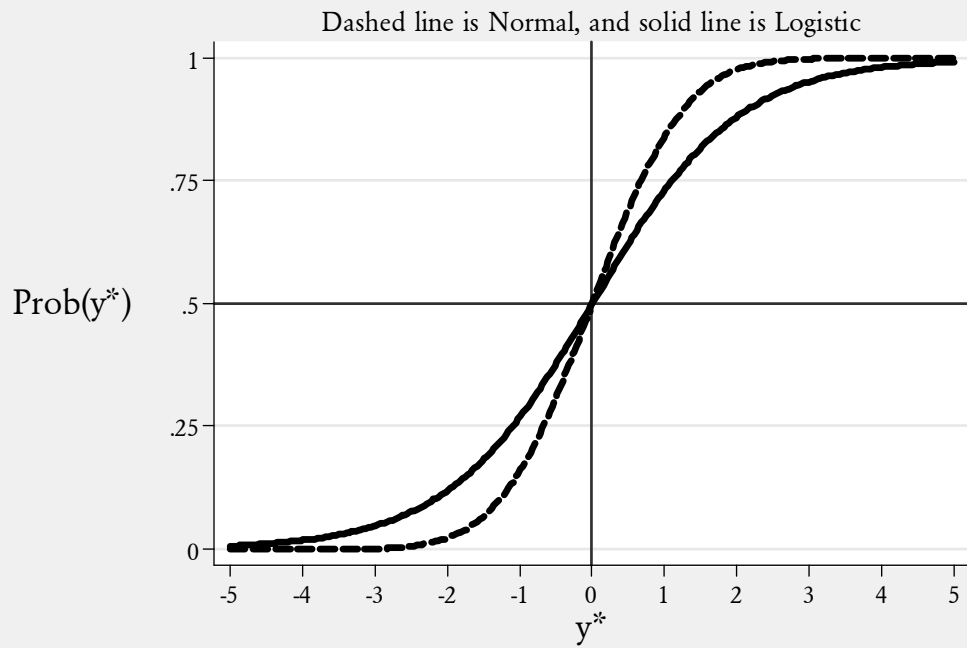


Figure 2: Distribution of Risk Attitudes Under EUT

Estimated with $N=158$ subjects from Harrison & Rutstrom [2005]
Power utility function: $r < 1$ is risk averse, $r = 1$ is risk neutral, $r > 1$ is risk loving

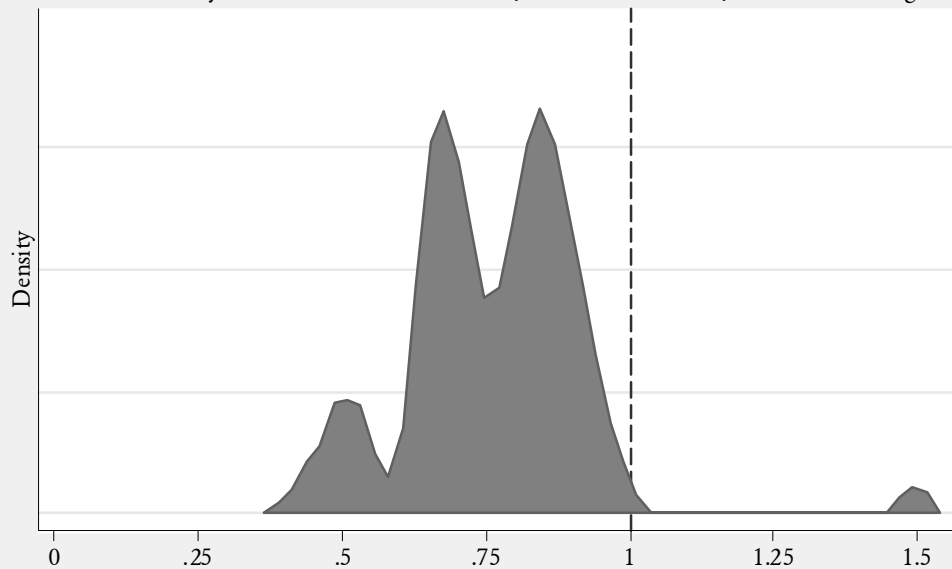


Table 1: Maximum Likelihood Estimates of Various Structural Models in Simple Binary Choice Lottery Experiments

Maximum likelihood estimation of EUT, RDU and CPT models.
N=9,311, based on 158 subjects from Harrison and Rutström [2009].

| Variable | Coefficient | Estimate | Standard Error | <i>p</i> -value | Lower 95% Confidence Interval | Upper 95% Confidence Interval |
|---|----------------|----------|----------------|-----------------|-------------------------------|-------------------------------|
| <i>A. EUT: Homogeneous Preferences Specification with No Stochastic Error</i> (log-likelihood = -6085.1) | | | | | | |
| r | Constant | 0.777 | 0.024 | <0.001 | 0.729 | 0.825 |
| <i>B. EUT: Homogeneous Preferences Specification</i> (log-likelihood = -6085.0) | | | | | | |
| r | Constant | 0.771 | 0.044 | <0.001 | 0.684 | 0.857 |
| μ | Constant | 0.966 | 0.134 | <0.001 | 0.704 | 1.228 |
| <i>C. EUT: Heterogeneous Preferences Specification</i> (log-likelihood = -5925.8) | | | | | | |
| r | Constant | 0.246 | 0.379 | 0.516 | -0.497 | 0.989 |
| | Female | -0.156 | 0.085 | 0.067 | -0.324 | 0.011 |
| | Black | -0.100 | 0.155 | 0.517 | -0.404 | 0.204 |
| | Hispanic | -0.232 | 0.137 | 0.089 | -0.500 | 0.036 |
| | Age in years | 0.030 | 0.019 | 0.103 | -0.006 | 0.067 |
| | Business major | -0.025 | 0.094 | 0.787 | -0.209 | 0.159 |
| | Low GPA | 0.061 | 0.099 | 0.534 | -0.133 | 0.256 |
| μ | Constant | 1.235 | 0.682 | 0.070 | -0.101 | 2.572 |
| | Female | -0.246 | 0.272 | 0.365 | -0.779 | 0.287 |
| | Black | -0.022 | 0.377 | 0.953 | -0.762 | 0.718 |
| | Hispanic | -0.020 | 0.421 | 0.962 | -0.846 | 0.806 |
| | Age in years | -0.018 | 0.028 | 0.531 | -0.073 | 0.037 |
| | Business major | 0.048 | 0.335 | 0.886 | -0.610 | 0.706 |
| | Low GPA | 0.335 | 0.385 | 0.384 | -0.419 | 1.088 |
| <i>D. RDU: Homogeneous Preferences Specification</i> (log-likelihood = -6083.4) | | | | | | |
| ρ | Constant | 0.763 | 0.043 | <0.001 | 0.678 | 0.848 |
| γ | Constant | 0.986 | 0.008 | <0.001 | 0.971 | 1.002 |
| μ | Constant | 0.938 | 0.128 | <0.001 | 0.686 | 1.190 |
| <i>E. CPT: Homogeneous Preferences Specification, Unconstrained Utility Functions</i> (log-likelihood = -5821.3) | | | | | | |
| α | Constant | 0.535 | 0.043 | <0.001 | 0.451 | 0.619 |
| β | Constant | 0.930 | 0.049 | <0.001 | 0.835 | 1.026 |
| λ | Constant | 0.338 | 0.081 | <0.001 | 0.179 | 0.497 |
| γ | Constant | 0.928 | 0.026 | <0.001 | 0.876 | 0.979 |
| μ | Constant | 0.557 | 0.075 | <0.001 | 0.410 | 0.704 |
| <i>F. CPT: Homogeneous Preferences Specification, Utility Functions Constrained to α=β</i> (log-likelihood = -5865.6) | | | | | | |
| α = β | Constant | 0.582 | 0.055 | <0.001 | 0.474 | 0.690 |
| λ | Constant | 0.693 | 0.160 | <0.001 | 0.778 | 1.007 |
| γ | Constant | 0.934 | 0.038 | <0.001 | 0.860 | 1.010 |
| μ | Constant | 0.743 | 0.139 | <0.001 | 0.471 | 1.015 |

Notes: All coefficients on variables refer to binary dummies, apart from age which is measured in years (average age in sample is 21). Low GPA is any reported GPA below 3.25.

Figure 3: Estimates of the Structural CPT Model With a Range of Assumed Reference Points

Estimated with N=158 subjects from Harrison & Rutstrom [2005]
N=156 subjects: 63 gain frame, 57 loss frame, and 36 mixed frame

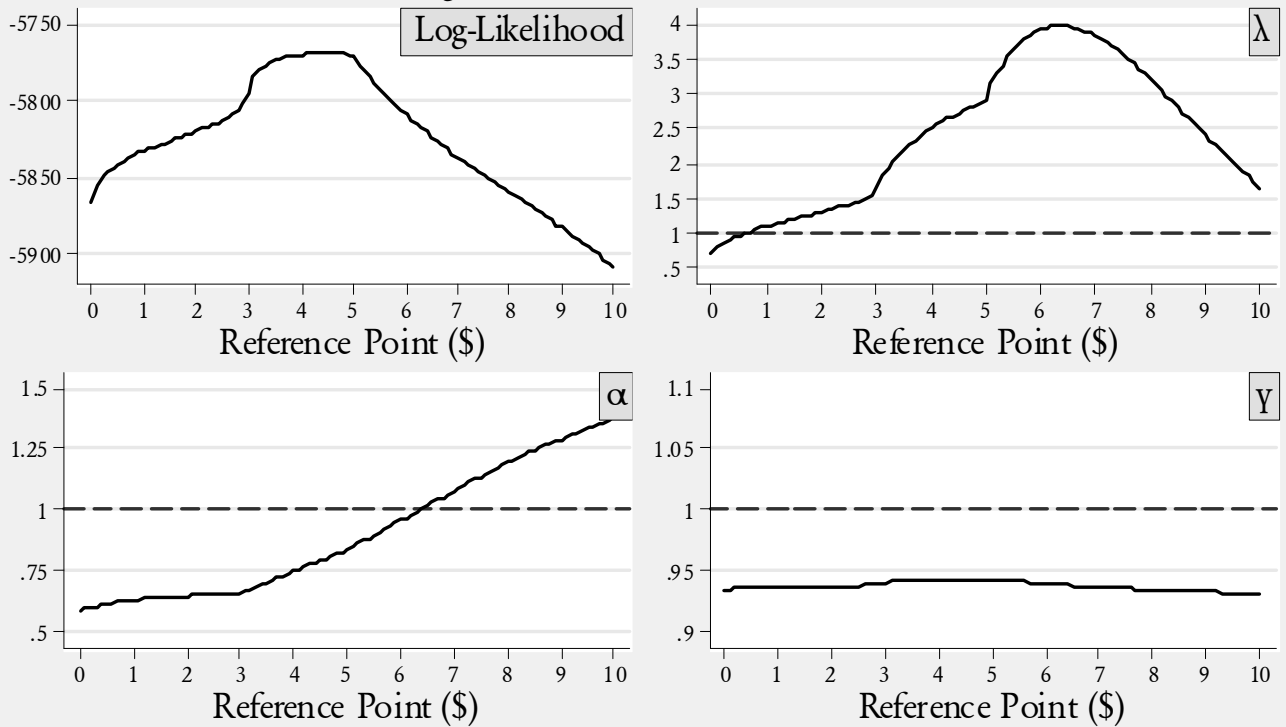


Figure 4: Opening Screen Shot for Laboratory Experiment



Figure 5: Prize Distribution and Display for Laboratory Experiment

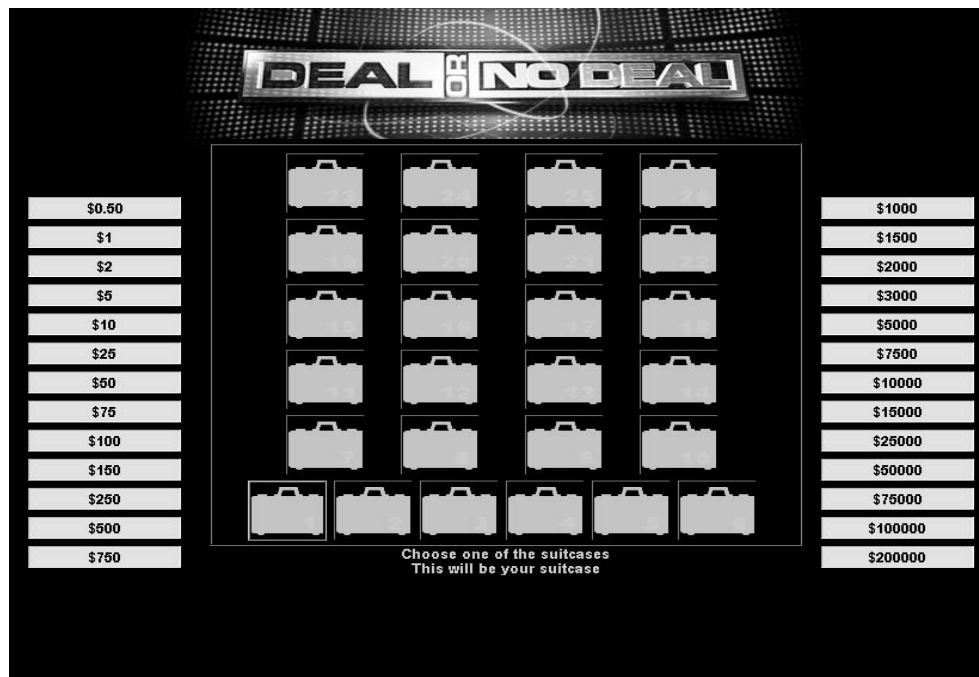


Figure 6: Typical Bank Offer in Laboratory Experiment

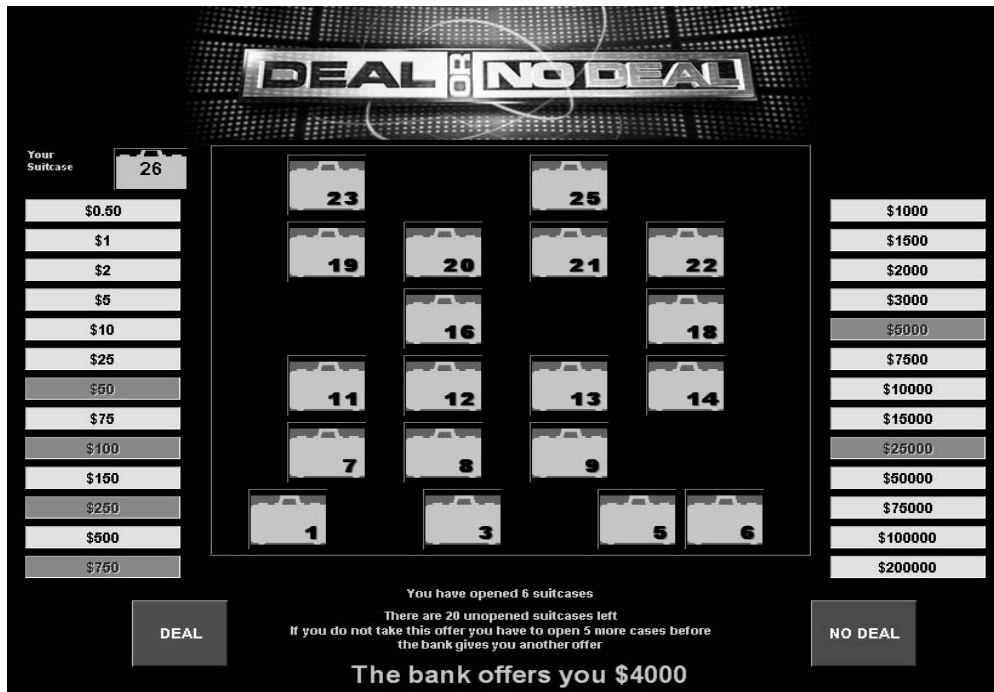


Figure 7: Information on Bank Offers in Laboratory Experiment

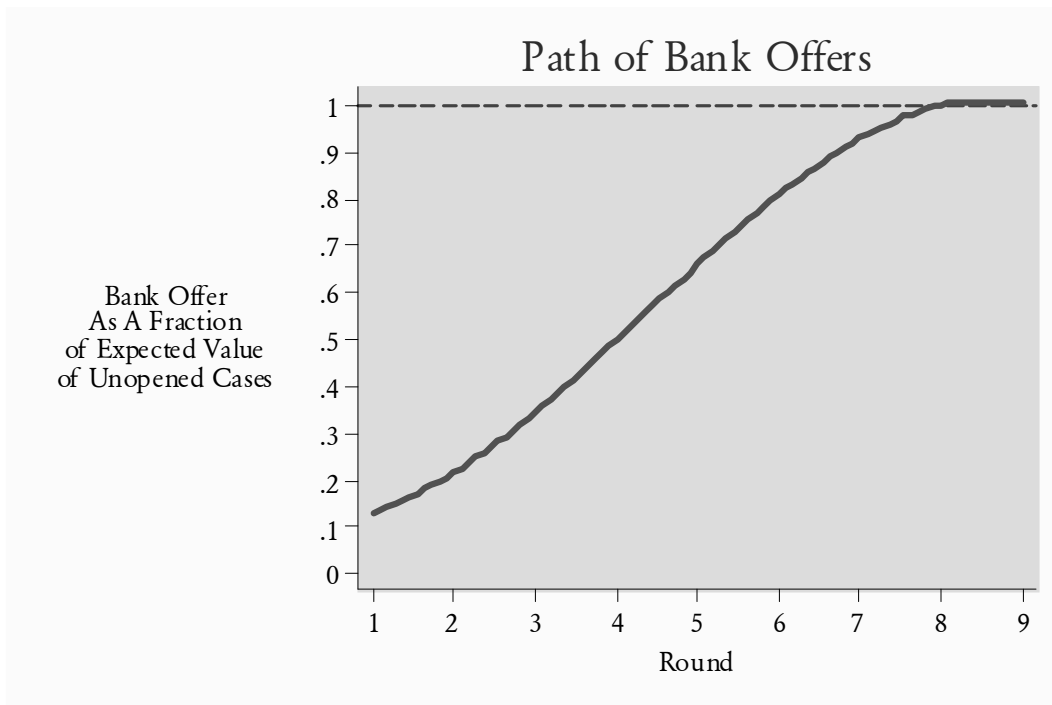
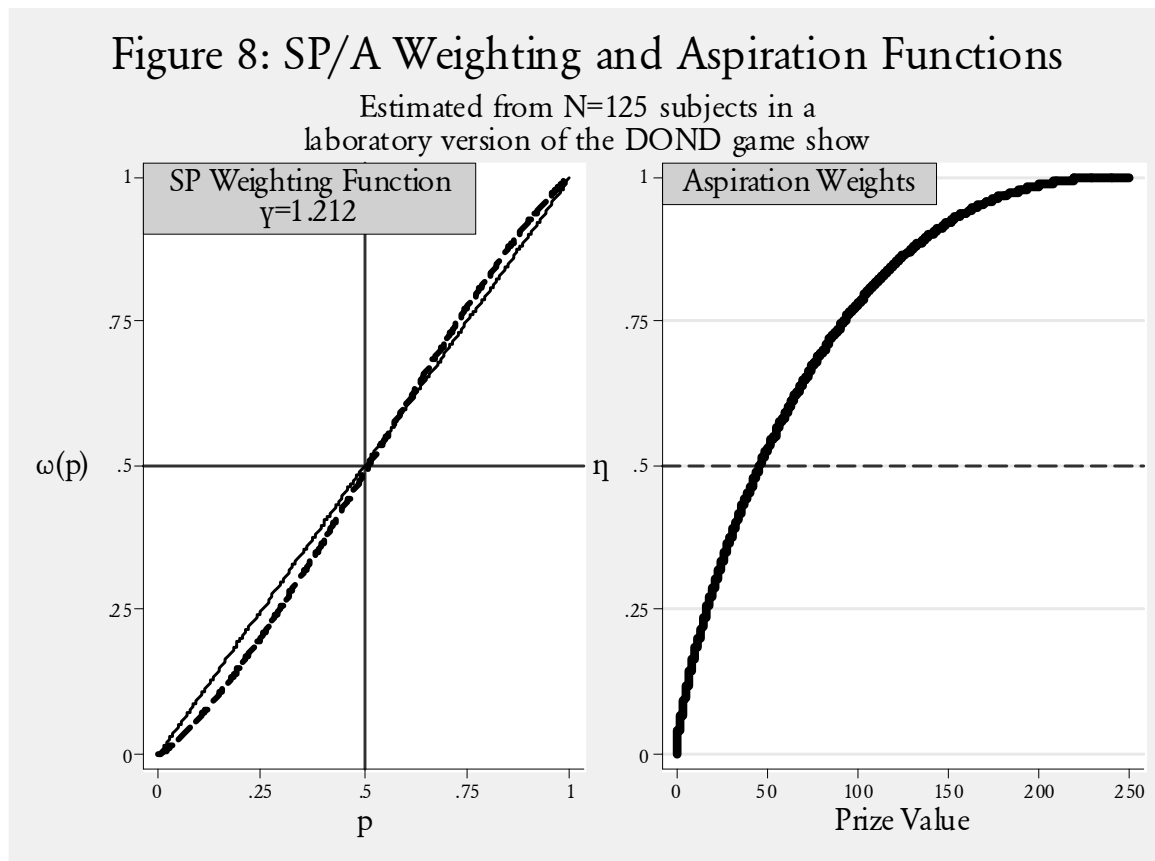


Table 2: Estimates for *Deal or No Deal* Laboratory Experiment Assuming SP/A

| Parameter | Estimate | Standard Error | Lower 95% Confidence Interval | Upper 95% Confidence Interval |
|------------------------------------|----------|----------------|-------------------------------|-------------------------------|
| ρ | 0.552 | 0.045 | 0.463 | 0.640 |
| γ | 1.211 | 0.017 | 1.178 | 1.245 |
| χ | 0.764 | 0.062 | 0.642 | 0.887 |
| ξ | 2.388 | 0.607 | 1.199 | 3.577 |
| Ψ ($\times 1000$) | 0.013 | 0.005 | 0.002 | 0.024 |
| μ | 0.082 | 0.009 | 0.063 | 0.010 |
| π^{SP} | 0.073 | 0.098 | -0.119‡ | 0.264 |
| $\pi^{\text{A}}=1-\pi^{\text{SP}}$ | 0.929 | 0.052 | 0.736 | 1.119‡ |

Note: ‡ Standard errors and confidence intervals calculated using the “delta method,” an approximation that is asymptotically valid, so estimated bounds of probabilities can lie outside the unit interval.



References

- Allais, Maurice, "The Foundations of Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School," in M. Allais & O. Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox* (Dordrecht, the Netherlands: Reidel, 1979).
- Andersen, Steffen; Harrison, Glenn W., Lau, Morten I., and Rutström, E. Elisabet, "Dual Criteria Decisions," *Working Paper 06-11*, Department of Economics, College of Business Administration, University of Central Florida, 2006.
- 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: Emerald, Research in Experimental Economics, Volume 12, 2008b).
- Andersen, Steffen; Harrison, Glenn W., and Rutström, E. Elisabet, "Choice Behavior, Asset Integration and Natural Reference Points," *Working Paper 06-04*, Department of Economics, College of Business Administration, University of Central Florida, 2006.
- Atkinson, A.C., "A Method for Discriminating Between Models," *Journal of the Royal Statistical Society, Series B*, 32, 1970, 323-344.
- Ballinger, T. Parker, and Wilcox, Nathaniel T., "Decisions, Error and Heterogeneity," *Economic Journal*, 107, July 1997, 1090-1105.
- Baumol, William J., "An Expected Gain-Confidence Limit Criterion for Portfolio Selection," *Management Science*, 10, 1963, 174-182.
- Becker, Gordon M.; DeGroot, Morris H., and Marschak, Jacob, "Stochastic Models of Choice Behavior," *Behavioral Science*, 8, 1963, 41-55.
- Benartzi, Shlomo, and Thaler, Richard H., "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics*, 111(1), February 1995, 75-92.
- Birnbaum, Michael H., "Causes of Allais Common Consequence Paradoxes: An Experimental Dissection," *Journal of Mathematical Psychology*, 48, 2004a, 87-106.
- Birnbaum, Michael H., "Tests of Rank-Dependent Utility and Cumulative Prospect Theory in Gambles Represented by Natural Frequencies: Effects of Format, Event Framing, and Branch Splitting," *Organizational Behavior and Human Decision Processes*, 95, 2004b, 40-65.
- Birnbaum, Michael H., "Evidence Against Prospect Theories in Gambles with Positive, Negative, and Mixed Consequences," *Journal of Economic Psychology*, 27, 2006, 737-761.
- Birnbaum, Michael H., "Evaluation of the Priority Heuristic as a Descriptive Model of Risky Decision Making: Comment on Brandstätter, Gigerenzer, and Hertwig (2006)," *Psychological Review*, 115(1), January 2008, 253-260.
- Birnbaum, Michael H., and Chavez, Alfredo, "Tests of Theories of Decision-Making: Violations of Branch Independence and Distribution Independence," *Organizational Behavior and Human Decision Processes*, 71(2), August 1997, 161-194.

- Birnbaum, Michael H.; Coffey, Gregory; Mellers, Barbara A., and Weiss, Robin, "Utility Measurement: Configural-Weight Theory and the Judge's Point of View," *Journal of Experimental Psychology: Human Perception & Performance*, 18(2), 1992, 331-346.
- Birnbaum, Michael H., and McIntosh, William Ross, "Violations of Branch Independence in Choices Between Gambles," *Organizational Behavior and Human Decision Processes*, 67(1), July 1996, 91-110.
- Birnbaum, Michael H., and Navarrete, Juan B., "Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence," *Journal of Risk and Uncertainty*, 17, 1998, 17-49.
- Brandstätter, Eduard; Gigerenzer, Gerd, and Hertwig, Ralph, "The Priority Heuristic: Making Choices Without Trade-Offs," *Psychological Review*, 113(2), 2006, 409-432.
- Brunswik, Egon, "Representative design and probabilistic theory in a functional psychology," *Psychological Review*, 62, 1955, 193-217.
- Burke, Michael S.; Carter, John R.; Gominiak, Robert D., and Ohl, Daniel F., "An Experimental Note on the Allais Paradox and Monetary Incentives," *Empirical Economics*, 21, 1996, 617-232.
- Camerer, Colin; Babcock, Linda; Loewenstein, George, and Thaler, Richard, "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 112, May 1997, 407-441.
- Camerer, Colin F., "Three Cheers – Psychological, Theoretical, Empirical – for Loss Aversion," *Journal of Marketing Research*, XLII, May 2005, 129-133.
- Camerer, Colin F., and Ho, Teck-Hua, "Violations of the Betweenness Axiom and Nonlinearity in Probability," *Journal of Risk and Uncertainty*, 8, 1994, 167-196.
- Cameron, A. Colin, and Trivedi, Pravin K., *Microeconometrics: Methods and Applications* (New York: Cambridge University Press, 2005).
- Clarke, Kevin A., "Nonparametric Model Discrimination in International Relations," *Journal of Conflict Resolution*, 47(1), February 2003, 72-93.
- Clarke, Kevin A., "A Simple Distribution-Free Test for Non-Nested Model Selection," *Political Analysis*, 15(3), 2007, 347-363.
- Conlisk, John, "Three Variants on the Allais Example," *American Economic Review*, 79(3), June 1989, 392-407.
- Cox, David R., "Tests of Separate Families of Hypotheses," in E.G. Charatsis (ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley: University of California Press, Volume 1, 1961, 105-123).
- Cox, David R., "Further Results on Tests of Separate Families of Hypotheses," *Journal of the Royal Statistical Society, Series B*, 24, 1962, 406-424.
- Davidson, Russell, and MacKinnon, James G., *Estimation and Inference in Econometrics* (New York: Oxford University Press, 1993).

- Edwards, Ward, "Subjective Probabilities Inferred from Decisions," *Psychological Review*, 69, 1962, 109-135.
- El-Gamal, Mahmoud A., and Grether, David M., "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 432, December 1995, 1137-1145.
- Ericsson, K. Anders, and Simon, Herbert A., *Protocol Analysis: Verbal Reports as Data* (Cambridge, MA: MIT Press, 1993, Revised Edition).
- Fan, Chinn-Ping, "Allais Paradox in the Small," *Journal of Economic Behavior & Organization*, 49, 2002, 411-421.
- Farber, Henry S., "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 113(1), 2005, 46-82.
- Fechner, Gustav, *Elements of Psychophysics* (New York: Holt, Rinehart & Winston, Volume 1, 1966; originally published 1860).
- Gigerenzer, Gerd, and Kurz., E., "Vicarious Functioning Reconsidered: A Fast and Frugal Lens Model," in K.R. Hammond and T.R. Stewart (eds.), *The Essential Brunswick: Beginnings, Explinations, Applications* (New York: Oxford University Press, 2001).
- Gollier, Christian, *The Economics of Risk and Time* (Cambridge, MA: MIT Press, 2001).
- Gonzalez, Richard, and Wu, George, "On the Shape of the Probability Weighting Function," *Cognitive Psychology*, 38, 1999, 129-166.
- Gul, Faruk, and Pesendorfer, Wolfgang, "The Case for Mindless Economics," in A. Caplin and A. Schotter (eds.), *Handbook of Economic Methodologies* (New York: Oxford University Press, 2007).
- Harless, David W., and Camerer, Colin F., "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica*, 62(6), November 1994, 1251-1289.
- Harrison, Glenn W., "Expected Utility Theory and The Experimentalists," *Empirical Economics*, 19(2), 1994, 223-253.
- Harrison, Glenn W., "Neuroeconomics: A Critical Reconsideration," *Economics and Philosophy*, 24, 2008, 203-244.
- Harrison, Glenn W.; Humphrey, Steven J., and Verschoor, Arjan, "Choice Under Uncertainty: Evidence from Ethiopia, India and Uganda," *The Economic Journal*, 120, March 2010, 80-104.
- Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95(3), June 2005, 897-901.
- Harrison, Glenn W.; Johnson, Eric; McInnes, Melayne M., and Rutström, E. Elisabet, "Measurement With Experimental Controls," in M. Boumans (ed.), *Measurement in Economics: A Handbook* (San Diego, CA: Elsevier, 2007).
- Harrison, Glenn W., and List, John A., "Field Experiments," *Journal of Economic Literature*, 42(4), December 2004, 1013-1059.

- 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).
- Harrison, Glenn W., and Rutström, E. Elisabet, "Expected Utility *And* Prospect Theory: One Wedding and a Decent Funeral," *Experimental Economics*, 12(2), 2009, 133-158.
- Hertwig, Ralph, and Ortmann, Andreas, "Experimental Practices in Economics: A Challenge for Psychologists?" *Behavioral and Brain Sciences*, 24, 2001, 383–451.
- Hertwig, Ralph, and Ortmann, Andreas, "The Cognitive Illusion Controversy: A Methodological Debate in Disguise that Matters to Economists," in R. Zwick & A. Rapoport (eds.), *Experimental Business Research, Volume III: Marketing, Accounting, and Cognitive Perspectives*, (New York: Springer Dordrecht, 2005).
- Hey, John, "Experimental Investigations of Errors in Decision Making Under Risk," *European Economic Review*, 39, 1995, 633-640.
- Hey, John D., "Experimental Economics and the Theory of Decision Making Under Uncertainty," *Geneva Papers on Risk and Insurance Theory*, 27(1), June 2002, 5-21.
- Hey, John D., and Orme, Chris, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), November 1994, 1291-1326.
- Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), December 2002, 1644-1655.
- Holt, Charles A., and Laury, Susan K., "Risk Aversion and Incentive Effects: New Data Without Order Effects," *American Economic Review*, 95(3), June 2005, 902-904.
- Hosmer, David W., and Lemeshow, Stanley, *Applied Logistic Regression* (New York: Wiley, Second Edition, 2000).
- Johnson, Eric J.; Schulte-Mecklenbeck, Michael, and Willemsen, Martijn C., "Process Models Deserve Process Data: A Comment on Brandstätter, Gigerenzer, and Hertwig (2006)," *Psychological Review*, 115(1), January 2008, 263-272.
- Johnson, Norman L.; Kotz, Samuel, and Balakrishnan, N., *Continuous Univariate Distributions, Volume 2* (New York: Wiley, Second Edition, 1995).
- Kahneman, Daniel, and Tversky, Amos, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 1979, 263-291.
- Knutson, Brian, and Peterson, Richard, "Neurally Reconstructing Expected Utility," *Games and Economic Behavior*, 52(2), August 2005, 305-315.
- Köbberling, Veronika, and Wakker, Peter P., "An Index of Loss Aversion," *Journal of Economic Theory*, 122, 2005, 119-131.
- Kőszegi, Botond, and Rabin, Matthew, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121(4), November 2006, 1133-1165.
- Kőszegi, Botond, and Rabin, Matthew, "Reference-Dependent Risk Attitudes," *American Economic Review*, 97(4), September 2007, 1047-1073.

- Leland, W. Jonathan, "Generalized Similarity Judgements: An Alternative Explanation for Choice Anomalies," *Journal of Risk & Uncertainty*, 9, 1994, 151-172.
- Liang, K-Y., and Zeger, S.L., "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 1986, 13-22.
- Loomes, Graham, "Modelling the Stochastic Component of Behavior in Experiments: Some Issues for the Interpretation of Data," *Experimental Economics*, 8, 2005, 301-323.
- Loomes, Graham; Moffatt, Peter G., and Sugden, Robert, "A Microeconomic Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24(2), 2002, 103-130.
- 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.
- Lopes, Lola L., "Risk and Distributional Inequality," *Journal of Experimental Psychology: Human Perception and Performance*, 10(4), August 1984, 465-484.
- Lopes, Lola L., "Algebra and Process in the Modeling of Risky Choice," in J.R. Busemeyer, R. Hastie & D.L. Medin (eds), *Decision Making from a Cognitive Perspective* (San Diego: Academic Press, 1995).
- Lopes, Lola L., and Oden, Gregg C., "The Role of Aspiration Level in Risky Choice: A Comparison of Cumulative Prospect Theory and SP/A Theory," *Journal of Mathematical Psychology*, 43, 1999, 286-313.
- Luce, R. Duncan, *Individual Choice Behavior* (New York: Wiley, 1959).
- Luce, R. Duncan, and Fishburn, Peter C., "Rank and Sign-Dependent Linear Utility Models for Finite First-Order Gambles," *Journal of Risk & Uncertainty*, 4, 1991, 29-59.
- Manski, Charles F., "The Maximum Score Estimator of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 1975, 205-228.
- Novemsky, Nathan, and Kahneman, Daniel, "The Boundaries of Loss Aversion," *Journal of Marketing Research*, XLII, May 2005, 119-128.
- Oden, Gregg C., and Lopes, Lola L., "Risky Choice With Fuzzy Criteria," *Psychologische Beiträge*, 39, 1997, 56-82.
- Oehlert, Gary W., "A Note on the Delta Method," *The American Statistician*, 46(1), February 1992, 27-29.
- Payne, John W.; Bettman, James R., and Johnson, Eric J., *The Adaptive Decision Maker* (New York: Cambridge University Press, 1993).
- Pesaran, M.H., "Pitfalls of Testing Non-Nested Hypotheses by the Lagrange Multiplier Method," *Journal of Econometrics*, 17, 1981, 323-331.
- Pollak, Robert A., and Wales, Terence J., "The Likelihood Dominance Criterion: A New Approach to Model Selection," *Journal of Econometrics*, 47, 1991, 227-242.

- Post, Thierry; van den Assem, Martijn J.; Baltussen, Guido, and Thaler, Richard H., "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 98(1), March 2008, 38-71.
- Prelec, Drazen, "The Probability Weighting Function," *Econometrica*, 66, 1998, 497-527.
- Quandt, Richard E., "A Comparison of Methods for Testing Nonnested Hypotheses," *Review of Economics and Statistics*, 56, 1974, 92-99.
- Quiggin, John, "A Theory of Anticipated Utility," *Journal of Economic Behavior & Organization*, 3(4), 1982, 323-343.
- Rieger, Marc Oliver and Wang, Mei, "Cumulative Prospect Theory and the St. Petersburg Paradox," *Economic Theory*, 28, 2006, 665-679.
- Rieger, Marc Oliver and Wang, Mei, "What is Behind the Priority Heuristic? A Mathematical Analysis," *Psychological Review*, 115(1), January 2008, 274-280.
- Rogers, W. H., "Regression standard errors in clustered samples," *Stata Technical Bulletin*, 13, 1993, 19-23.
- Roy, A.D., "Safety First and the Holding of Assets," *Econometrica*, 20(3), July 1952, 431-449.
- Roy, A.D., "Risk and Rank or Safety First Generalised," *Economica*, 23, August 1956, 214-228.
- Rubinstein, Ariel, "Similarity and Decision-making Under Risk (Is There a Utility Theory Resolution to the Allais Paradox?)," *Journal of Economic Theory*, 46, 1988, 145-153.
- 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.
- Samuelson, Paul A., *Foundations of Economic Analysis* (Boston: Harvard University Press, 1947).
- Schmidt, Ulrich; Starmer, Chris, and Sugden, Robert, "Third-Generation Prospect Theory," *Journal of Risk and Uncertainty*, 36(3), June 2008, 203-223.
- Smith, Vernon L., "Constructivist and Ecological Rationality in Economics," *American Economic Review*, 93(3), June 2003, 465-508.
- Starmer, Chris, "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk," *Journal of Economic Literature*, 38, June 2000, 332-382.
- Starmer, Chris, and Sugden, Robert, "Violations of the Independence Axiom in Common Ratio Problems: An Experimental Test of Some Competing Hypotheses," *Annals of Operational Research*, 19, 1989, 79-102.
- StataCorp, *Stata Statistical Software: Release 10* (College Station, TX: Stata Corporation, 2007).
- Thorngate, Warren, "Efficient Decision Heuristics," *Behavioral Science*, 25, 1980, 219-225.
- Tversky, Amos, "Intransitivity of Preferences," *Psychological Review*, 76, 1969, 31-48.
- Tversky, Amos, and Kahneman, Daniel, "Advances in Prospect Theory: Cumulative Representations of Uncertainty," *Journal of Risk & Uncertainty*, 5, 1992, 297-323.

- Vuong, Quang H., "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), March 1989, 307-333.
- Wakker, Peter P., "Transforming Probabilities Without Violating Stochastic Dominance," in E. Roskam (ed.), *Mathematical Psychology in Progress* (Berlin: Springer, 1989).
- Wilcox, Nathaniel T., "Predicting Individual Risky Choices Out-of-Context: A Critical Stochastic Modeling Primer and Monte Carlo Study," 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*, 2009 forthcoming, doi:10.1016/j.jeconom.2009.10.012.
- Williams, Rick L., "A Note on Robust Variance Estimation for Cluster-Correlated Data," *Biometrics*, 56, June 2000, 645-646.
- Wooldridge, Jeffrey, "Cluster-Sample Methods in Applied Econometrics," *American Economic Review (Papers & Proceedings)*, 93, May 2003, 133-138.
- Yaari, Menahem E., "The Dual Theory of Choice under Risk," *Econometrica*, 55(1), 1987, 95-115.

Appendix A: Experimental Procedures for Binary Lottery Choice Task

Subjects were presented with 60 lottery pairs, each represented as a “pie” showing the probability of each prize. Harrison and Rutström [2008; Figure 4] illustrates one such representation. The subject could choose the lottery on the left or the right, or explicitly express indifference (in which case the experimenter would flip a coin on the subject’s behalf). After all 60 lottery pairs were evaluated, three were selected at random for payment.⁴⁸ The lotteries were presented to the subjects in color on a private computer screen,⁴⁹ and all choices recorded by the computer program. This program also recorded the time taken to make each choice. In addition to the choice tasks, the subjects provided information on demographic and other personal characteristics.

In the gain frame experiments the prizes in each lottery were \$0, \$5, \$10 and \$15, and the probabilities of each prize varied from choice to choice, and from lottery to lottery. In the loss frame experiments subjects were given an initial endowment of \$15, and the corresponding prizes from the gain frame lotteries were transformed to be -\$15, -\$10, -\$5 and \$0. Hence the final outcomes, inclusive of the endowment, were the same in the gain frame and loss frame. In the mixed frame experiments subjects were given an initial endowment of \$8, and the prizes were transformed to be -\$8, -\$3, \$3 and \$8, generating final outcomes inclusive of the endowment of \$0, \$5, \$11 and \$16.⁵⁰

In addition to the fixed endowment, each subject received a random endowment between \$1 and \$10. This endowment was generated using a uniform distribution defined over whole dollar amounts, operationalized by a 10-sided die. The purpose of this random endowment is to test for endowment effects on the choices.

⁴⁸ The typical application of the random lottery incentive mechanism in experiments such as these would have one choice selected at random. We used three to ensure comparability of rewards with other experiments in which subjects made choices over 40 or 20 lotteries, and where 2 lotteries or 1 lottery was respectively selected at random to be played out.

⁴⁹ The computer laboratory used for these experiments has 28 subject stations. Each screen is “sunken” into the desk, and subjects were typically separated by several empty stations due to staggered recruitment procedures. No subject could see what the other subjects were doing, let alone mimic what they were doing since each subject was started individually at different times.

⁵⁰ These final outcomes differ by \$1 from the two highest outcomes for the gain frame and mixed frame, because we did not want to offer prizes in fractions of dollars.

The probabilities used in each lottery ranged roughly evenly over the unit interval. Values of 0, 0.13, 0.25, 0.37, 0.5, 0.62, 0.75 and 0.87 were used.⁵¹ The presentation of a given lottery on the left or the right was determined at random, so that the “left” or “right” lotteries did not systematically reflect greater risk or greater prize range than the other.

Subjects were recruited at the University of Central Florida, primarily from the College of Business Administration, using the online recruiting application at ExLab (<http://exlab.bus.ucf.edu>). Each subject received a \$5 fee for showing up to the experiments, and completed an informed consent form. Subjects were deliberately recruited for “staggered” starting times, so that the subject would not pace their responses by any other subject. Each subject was presented with the instructions individually, and taken through the practice sessions at an individual pace. Since the rolls of die were important to the implementation of the objects of choice, the experimenters took some time to give each subject “hands-on” experience with the (10-sided, 20-sided and 100-sided) die being used. Subjects were free to make their choices as quickly or as slowly as they wanted.

Our data consists of responses from 158 subjects making 9311 choices that do not involve indifference. Only 1.7% of the choices involved explicit choice of indifference, and to simplify we drop those in estimation unless otherwise noted. Of these 158 subjects, 63 participated in gain frame tasks, 37 participated in mixed frame tasks, and 58 participated in loss frame tasks.

⁵¹ To ensure that probabilities summed to one, we also used probabilities of 0.26 instead of 0.25, 0.38 instead of 0.37, 0.49 instead of 0.50 or 0.74 instead of 0.75.

Appendix B: Estimation Using Maximum Likelihood

(NOT FOR PUBLICATION)

Economists in a wide range of fields are now developing customized likelihood functions to correspond to specific models of decision-making processes. These demands derive partly from the need to consider a variety of parametric functional forms, but also because these models often specify non-standard decision rules that have to be “written out by hand.” Thus it is becoming common to see user-written maximum likelihood estimates, and less use of pre-packaged model specifications.

These pedagogic notes document the manner in which one can estimate maximum likelihood models of utility functions within *Stata*.⁵² However, we can quickly go beyond “utility functions” and consider a wide range of decision-making processes, to parallel the discussion in the text. We start with a standard CRRA utility function and binary choice data over two lotteries, assuming EUT. This step illustrates the basic economic and statistical logic, and introduces the core *Stata* syntax. We then quickly consider an extension to consider loss aversion and probability weighting from PT, the inclusion of “stochastic errors,” and the estimation of utility numbers themselves to avoid any parametric assumption about the utility function. Once the basic syntax is defined from the first example, it is possible to quickly jump to other likelihood functions using different data and specifications. Of course, this is just a reflection of the “extensible power” of a package such as *Stata*, once one understands the basic syntax.⁵³

⁵² The exposition is deliberately transparent to economists. Most of the exposition here would be redundant for those familiar with Gould, Pitblado and Sribney [2006] or even Rabe-Hesketh and Everitt [2004; ch.13]. It is easy to find expositions of maximum likelihood in *Stata* that are more general and elegant for their purpose, but for those trying to learn the basic tools for the first time that elegance can just appear to be needlessly cryptic coding, and actually act as an impediment to comprehension. There are good reasons that one wants to build more flexible and computationally efficient models, but ease of comprehension is rarely one of them. StataCorp [2007] documents the latest version 10 of *Stata*, but the exposition of the maximum likelihood syntax is minimal in that otherwise extensive documentation.

⁵³ Paarsch and Hong [2006; Appendix A.8] provide a comparable introduction to the use of MATLAB for estimation of structural models of auctions. Unfortunately their documentation contains no “real data” to evaluate the programs on.

B1. Estimating a CRRA Utility Function

Consider the simple CRRA specification in §1.1. This is an EUT model, with a CRRA utility function, and no stochastic error specification. The following *Stata* program defines the model, in this case using the lottery choices of Harrison and Rutström [2009], which are a replication of the experimental tasks of Hey and Orme [1994] as described in Appendix A:

```
* define Original Recipe EUT with CRRA and no errors
program define ML_eut0

    args lnf r
    tempvar prob01 prob11 prob21 prob31 prob0r prob1r prob2r prob3r y0 y1 y2 y3
    tempvar euL euR euDiff euRatio tmp lnf_eut lnf_pt p1 p2 f1 f2

    quietly {

        * construct likelihood for EUT
        generate double `prob01' = $ML_y2
        generate double `prob11' = $ML_y3
        generate double `prob21' = $ML_y4
        generate double `prob31' = $ML_y5

        generate double `prob0r' = $ML_y6
        generate double `prob1r' = $ML_y7
        generate double `prob2r' = $ML_y8
        generate double `prob3r' = $ML_y9

        generate double `y0' = ($ML_y14+$ML_y10)^`r'
        generate double `y1' = ($ML_y14+$ML_y11)^`r'
        generate double `y2' = ($ML_y14+$ML_y12)^`r'
        generate double `y3' = ($ML_y14+$ML_y13)^`r'

        gen double `euL' = (`prob01'*`y0')+(`prob11'*`y1')+(`prob21'*`y2')+(`prob31'*`y3')
        gen double `euR' = (`prob0r'*`y0')+(`prob1r'*`y1')+(`prob2r'*`y2')+(`prob3r'*`y3')

        generate double `euDiff' = `euR' - `euL'

        replace `lnf' = ln(normal(`euDiff')) if $ML_y1==1
        replace `lnf' = ln(normal(-`euDiff')) if $ML_y1==0

    }
end
```

This program makes more sense when one sees the command line invoking it, and supplying it with values for all variables. The simplest case is where there are no explanatory variables for the CRRA coefficient (we cover those below):

```
ml model lf ML_eut0 (r: Choices P0left P1left P2left P3left P0right P1right P2right
P3right prize0 prize1 prize2 prize3 stake = ) if Choices~., cluster(id)
technique(nr) maximize
```

The “ml model” part invokes the *Stata* maximum likelihood model specification routine, which essentially reads in the ML_eut0 program defined above and makes sure that it does not violate any syntax rules. The “lf” part of “lf ML_eut0” tells this routine that this is a particular type of likelihood specification (specifically, that the routine ML_eut0 does not calculate analytical derivatives, so those must be calculated numerically). The part in brackets defines the equation for

the CRRA coefficient r . The “r:” part just labels this equation, for output display purposes and to help reference initial values if they are specified for recalcitrant models. There is no need for the “r:” here to match the “r” inside the ML_eut0 program; we could have referred to “rEUT:” in the “ml model” command. We use the same “r” to help see the connection, but it is not essential.

The “Choices p0left p1left p2left p3left p0right p1right p2right p3right prize0 prize1 prize2 prize3 stake” part tells the program what observed values and data to use. This allows one to pass parameter values as well as data to the likelihood evaluator defined in ML_eut0. Each item in this list translates into a $\$ML_y^*$ variable referenced in the ML_eut0 program, where $*$ denotes the order in which it appears in this list. Thus the data in variable Choices, which consists of 0's and 1's for choices (and a dot, to signify “missing”), is passed to the ML_eut0 program as variable $\$ML_y1$. Variable p0left, which holds the probabilities of the first prize of the lottery presented to subjects on the left of their screen, is passed as $\$ML_y2$, and so on. Finally, variable stake, holding the values of the initial endowments provided to subjects, gets passed as variable $\$ML_y14$. It is good programming practice to then define these in some less cryptic manner, as we do just after the “quietly” line in ML_eut0. This does not significantly slow down execution, and helps avoid cryptic code. There is no error if some variable that is passed to ML_eut0 is not referenced in ML_eut0.

Once the data is passed to ML_eut0 the likelihood function can be evaluated. By default, it assumes a constant term, so when we have “=)” in the above command line, this is saying that there are no other explanatory variables. We add some below, but for now this model is just assuming that one CRRA coefficient characterizes all choices by all subjects. That is, it assumes that everyone has the same risk preference.

We restrict the data that is passed to only include strict preferences, hence the “if Choices~=” part at the end of the command line. The response of indifference was allowed in this experiment, and we code it as a “missing” value. Thus the estimation only applies to the sub-sample of strict preferences. One could modify the likelihood function to handle indifference.

Returning to the `ML_eut0` program, the “args” line defines some arguments for this program. When it is called, by the default Newton-Raphson optimization routine within *Stata*, it accepts arguments in the “r” array and returns a value for the log-likelihood in the “lnf” scalar. In this case “r” is the vector of coefficient values being evaluated.

The “tempvar” lines create temporary variables for use in the program. These are temporary in the sense that they are only local to this program, and hence can be the same as variables in the main calling program. Once defined they are referred to with the `ML_eut0` program by adding the funny left single-quote mark ``` and the regular right single-quote mark `'`. Thus temporary variable `euL`, to hold the expected utility of the left lottery, is referred to as ``euL'` in the program.⁵⁴

The “quietly” line defines a block of code that is to be processed without the display of messages. This avoids needless display of warning messages, such as when some evaluation returns a missing value. Errors are not skipped, just display messages.⁵⁵

The remaining lines should make sense to any economist from the comment statements. The program simply builds up the expected utility of each lottery, using the CRRA specification for the utility of the prizes. Then it uses the probit index function to define the likelihood values. The actual responses, stored in variable `Choices` (which is internal variable `$ML_y1`), are used at the very end to define which side of the probit index function this choice happens to be. The logit index specification is just as easy to code up: you replace “normal” with “invlogit” and you are done! The most important feature of this specification is that one can “build up” the latent index with as many programming lines as needed. Thus, as illustrated below, it is an easy matter to write out more detailed models, such as required for estimation of PT specifications or mixture models.

The “cluster(id)” command at the end tells *Stata* to treat the residuals from the same person as potentially correlated. It then corrects for this fact when calculating standard errors of estimates.

⁵⁴ Note that this is ``euL'` and not `'euL'`: beginning *Stata* users make this mistake a lot.

⁵⁵ Since the `ML_eut0` program is called many, many times to evaluate Jacobians and the like, these warning messages can clutter the screen display needlessly. During debugging, however, one normally likes to have things displayed, so the command “quietly” would be changed to “noisily” for debugging. Actually, we use the “ml check” option for debugging, as explained later, and never change this to “noisily.” Or we can display one line by using the “noisily” option, to debug specific calculations.

Invoking the above command line, with the “, maximize” option at the end to tell *Stata* to actually proceed with the optimization, generates this output:

```

initial:      log pseudolikelihood = -6453.8934
alternative:  log pseudolikelihood = -6351.1167
rescale:     log pseudolikelihood = -6351.1167
Iteration 0:  log pseudolikelihood = -6351.1167
Iteration 1:  log pseudolikelihood = -6086.8651 (backed up)
Iteration 2:  log pseudolikelihood = -6085.2315
Iteration 3:  log pseudolikelihood = -6085.1085
Iteration 4:  log pseudolikelihood = -6085.1078
Iteration 5:  log pseudolikelihood = -6085.1078

. ml display

                                Number of obs =          9311
                                Wald chi2(0)   =              .
                                Prob > chi2    =              .

Log pseudolikelihood = -6085.1078

                                (Std. Err. adjusted for 158 clusters in id)
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
                                |          Coef.   Robust          z   P>|z|   [95% Conf. Interval]
                                |          |          Std. Err.          |          |          |
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
                                | _cons |   .7768509   .0244973   31.71   0.000   .728837   .8248648
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

So we see that the optimization routine converged nicely, with no error messages or warnings about numerical irregularities at the end. The interim warning message is nothing to worry about: only worry if there is an error message of any kind at the end of the iterations. (Of course, lots of error message, particularly about derivatives being hard to calculate, usually flag convergence problems). The “ml display” command allows us to view the standard output, and is given after the “ml model” command. For our purposes the critical thing is the “_cons” line, which displays the maximum-likelihood estimate and it’s standard error. Thus we have estimated that $\hat{\alpha} = 0.777$. Since we employ the power CRRA function the maximum likelihood CRRA coefficient is $(1-\hat{\alpha}) = 0.223$ in this case. This indicates that these subjects are risk averse.

Before your program runs nicely it may have some syntax errors. The easiest way to check these is to issue the command

```
ml model lf ML_eut0 (r: Choices P0left P1left P2left P3left P0right P1right P2right
P3right prize0 prize1 prize2 prize3 stake = )
```

which is the same as before except that it drops off the material after the comma, which tells *Stata* to maximize the likelihood and how to handle the errors. This command simply tells *Stata* to read in the model and be ready to process it, but not to begin processing it. You would then issue the command

ml check

and *Stata* will provide some diagnostics. These are extremely informative if you use them, particularly for syntax errors.

The power of this approach becomes evident when we allow the CRRA coefficient to be determined by individual or treatment characteristics. To illustrate, consider the effect of allowing the CRRA coefficient to differ depending on the individual demographic characteristics of the subject, as explained in the text. Here is a list and sample statistics:

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|-----|-----|
| Female | 261 | .4444444 | .4978587 | 0 | 1 |
| Black | 261 | .091954 | .2895162 | 0 | 1 |
| Hispanic | 261 | .1302682 | .337245 | 0 | 1 |
| Age | 261 | 20.4751 | 3.695677 | 17 | 47 |
| Business | 261 | .4597701 | .4993364 | 0 | 1 |
| GPAlow | 261 | .4750958 | .5003388 | 0 | 1 |

The earlier command line is changed slightly at the “=” part to read “= Female Black Hispanic Age Business GPAlow)”, and no changes are made to ML_eut0. The results are as follows:

```
ml model lf ML_eut0 (r: Choices P0left P1left P2left P3left P0right P1right P2right
P3right prize0 prize1 prize2 prize3 stake = Female Black Hispanic Age Business
GPAlow), cluster(id) maximize
```

```
. ml model lf ML_eut0 (r: Choices P0left P1left P2left P3left P0right P1right P2right
P3right prize0 prize1 prize2 prize3 stake = $demog ), cluster(id) technique(bfgs)
difficult maximize continue
```

```
. ml display
```

```
Log pseudolikelihood = -9311.596
```

| | | |
|---------------|---|--------|
| Number of obs | = | 14718 |
| Wald chi2(6) | = | 53.90 |
| Prob > chi2 | = | 0.0000 |

(Std. Err. adjusted for 240 clusters in id)

| | Coef. | Robust Std. Err. | z | P> z | [95% Conf. Interval] | |
|----------|-----------|------------------|-------|-------|----------------------|-----------|
| Female | -.084377 | .0348794 | -2.42 | 0.016 | -.1527393 | -.0160146 |
| Black | -.1492832 | .0678229 | -2.20 | 0.028 | -.2822137 | -.0163527 |
| Hispanic | -.1949104 | .0765511 | -2.55 | 0.011 | -.3449478 | -.0448729 |
| Age | .0193883 | .0035912 | 5.40 | 0.000 | .0123498 | .0264268 |
| Business | -.0125927 | .0330864 | -0.38 | 0.703 | -.0774409 | .0522555 |
| GPAlow | .03372 | .030945 | 1.09 | 0.276 | -.026931 | .094371 |
| _cons | .4536579 | .0811472 | 5.59 | 0.000 | .2946124 | .6127035 |

So we see that the r parameter changes from $r=0.777$ to $r=0.454 - 0.084 \times \text{Female} - 0.149 \times \text{Black} \dots$ and so on. We can quickly find out what the average value of r is when we evaluate this model using the actual characteristics of each subject and the estimated coefficients:

```
. . predictnl r=xb(r)
. summ r if task==1
```

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|----------|-----|----------|-----------|----------|----------|
| r | 261 | .7842469 | .1169435 | .5107672 | 1.267938 |

So the average value is 0.784, extremely close to the earlier estimate of 0.777. Thus all we have done is provided a richer characterization of risk attitudes around roughly the same mean.

B2. Loss Aversion and Probability Weighting

It is a simple matter to specify different economic models. Two of the major structural features of Prospect Theory are probability weighting and loss aversion. The code below implements each of these specifications, using the parametric forms of Tversky and Kahneman [1992]. For simplicity we assume that the decision weights are the probability weights, and do not implement the rank-dependent transformation of probability weights into decision weights. Thus the model is strictly an implementation of Original Prospect Theory from Kahneman and Tversky [1979]. The extension to rank-dependent decision weights is messy from a programming perspective, and nothing is gained pedagogically here by showing it; Harrison [2006] shows the mess in full. Nevertheless, in our illustrations all choice options are binary in which case the probability weights and the decision weights are the same.

Note how much of this code is similar to ML_eut0, and the differences:

```
* define OPT specification with no errors
program define MLkt0

    args lnf alpha beta lambda gamma

    tempvar prob01 prob11 prob21 prob31 prob0r prob1r prob2r prob3r y0 y1 y2 y3
    tempvar euL euR euDiff euRatio tmp

    quietly {

        gen double `tmp' = (($ML_y2^`gamma')+($ML_y3^`gamma')+($ML_y4^`gamma')+($ML_y5^`gamma'))
        replace `tmp' = `tmp'^(1/`gamma')
        generate double `prob01' = ($ML_y2^`gamma')/`tmp'
        generate double `prob11' = ($ML_y3^`gamma')/`tmp'
        generate double `prob21' = ($ML_y4^`gamma')/`tmp'
        generate double `prob31' = ($ML_y5^`gamma')/`tmp'

        replace `tmp' = (($ML_y6^`gamma')+($ML_y7^`gamma')+($ML_y8^`gamma')+($ML_y9^`gamma'))
        replace `tmp' = `tmp'^(1/`gamma')
        generate double `prob0r' = ($ML_y6^`gamma')/`tmp'
        generate double `prob1r' = ($ML_y7^`gamma')/`tmp'
        generate double `prob2r' = ($ML_y8^`gamma')/`tmp'
        generate double `prob3r' = ($ML_y9^`gamma')/`tmp'

        generate double `y0' = .
        replace `y0' = ($ML_y10)^(`alpha') if $ML_y10>=0
    }
end
```

```

replace `y0' = -`lambda'*(-$ML_y10)^(`beta') if $ML_y10<0

generate double `y1' = .
replace `y1' = ($ML_y11)^(`alpha') if $ML_y11>=0
replace `y1' = -`lambda'*(-$ML_y11)^(`beta') if $ML_y11<0

generate double `y2' = .
replace `y2' = ($ML_y12)^(`alpha') if $ML_y12>=0
replace `y2' = -`lambda'*(-$ML_y12)^(`beta') if $ML_y12<0

generate double `y3' = .
replace `y3' = ($ML_y13)^(`alpha') if $ML_y13>=0
replace `y3' = -`lambda'*(-$ML_y13)^(`beta') if $ML_y13<0

gen double `euL'=(`prob01'*`y0')+(`prob11'*`y1')+(`prob21'*`y2')+(`prob31'*`y3')
gen double `euR'=(`prob0r'*`y0')+(`prob1r'*`y1')+(`prob2r'*`y2')+(`prob3r'*`y3')

generate double `euDiff' = `euR' - `euL'

replace `lnf' = ln(normal(`euDiff')) if $ML_y1==1
replace `lnf' = ln(normal(-`euDiff')) if $ML_y1==0

end
}

```

The first thing to notice is that the initial line “args lnf alpha beta lambda gamma” has more parameters than with ML_0. The “lnf” parameter is the same, since it is the one used to return the value of the likelihood function for trial values of the other parameters. But we now have four parameters instead of just one.

When we estimate this model we get this output:

```

. ml model lf MLkt0 (alpha: Choices P0left P1left P2left P3left P0right P1right
P2right P3right prize0 prize1 prize2 prize3 = ) (beta: ) (lambda: ) (gamma: ),
cluster(id) maximize

```

```

. ml display

```

```

Log pseudolikelihood = -9199.9139
Number of obs = 14718
Wald chi2(0) = .
Prob > chi2 = .

```

(Std. Err. adjusted for 240 clusters in id)

| | Coef. | Robust Std. Err. | z | P> z | [95% Conf. Interval] | |
|--------|----------|------------------|-------|-------|----------------------|----------|
| alpha | | | | | | |
| _cons | .6983474 | .0213405 | 32.72 | 0.000 | .6565207 | .740174 |
| beta | | | | | | |
| _cons | .7928385 | .0579259 | 13.69 | 0.000 | .6793058 | .9063711 |
| lambda | | | | | | |
| _cons | .6313889 | .0994708 | 6.35 | 0.000 | .4364297 | .826348 |
| gamma | | | | | | |
| _cons | .9194815 | .028761 | 31.97 | 0.000 | .8631111 | .975852 |

So we get estimates for all four parameters. *Stata* used the variable “_cons” for the constant, and since there are no characteristics here, that is the only variable to be estimated. We could also add demographic or other characteristics to any or all of these four parameters. We see that the utility curvature coefficients α and β are similar, and indicate concavity in the gain domain and convexity in the loss domain. Here we did not restrict these to be the same, as argued by Köbberling and Wakker [2005; §7], but that could easily be done. The loss aversion parameter λ is less than 1, which is a blow for PT since “loss aversion” calls for $\lambda > 1$. This could be due to the lack of restrictions on the α and β , but as is shown in the paper we estimate λ as less than 1 even under that restrictive assumption. And γ is very close to 1, which is the value that implies that $w(p) = p$ for all p , the EUT case. We can readily test some of these hypotheses:

```
. test [alpha]_cons=[beta]_cons
( 1)  [alpha]_cons - [beta]_cons = 0
      chi2( 1) =    2.64
      Prob > chi2 =    0.1043

. test [lambda]_cons=1
( 1)  [lambda]_cons = 1
      chi2( 1) =   13.73
      Prob > chi2 =    0.0002

. test [gamma]_cons=1
( 1)  [gamma]_cons = 1
      chi2( 1) =    7.84
      Prob > chi2 =    0.0051
```

So we see that PT is not doing so well here in relation to the *a priori* beliefs it comes packaged with, and that the deviation in λ is indeed statistically significant. Further γ is not significantly different from 1.

B3. Adding Stochastic Errors

In the text the Luce and Fechner “stochastic error stories” were explained. To add the Luce specification, popularized by HL, we return to base camp, the ML_eut0 program, and simply make two changes. We augment the arguments by one parameter, μ , to be estimated:


```
args lnf r mu
```

and then we revise the line defining the EU difference from

```
generate double `euDiff' = `euR' - `euL'
```

to

```
replace `euDiff' = (`euR'^(1/`mu'))/((`euR'^(1/`mu'))+(`euL'^(1/`mu')))
```

So this changes the latent preference index from being the difference to the ratio. But it is also adds the $1/\mu$ exponent to each expected utility. Apart from this change in the program, there is nothing extra that is needed. You just add one more parameter in the “ml model” stage, as we did for the PT extensions. In fact, HL cleverly exploit the fact that the latent preference index defined above is already in the form of a cumulative probability density function, since it ranges from 0 to 1, and is equal to $1/2$ when the subject is indifferent between the two lotteries. Thus, instead of defining the likelihood contribution by

```
replace `lnf' = ln(normal(`euDiff')) if $ML_y1==1  
replace `lnf' = ln(normal(-`euDiff')) if $ML_y1==0
```

we can use

```
replace `lnf' = ln(`euDiff') if $ML_y1==1  
replace `lnf' = ln(1-`euDiff') if $ML_y1==0
```

instead.

The Fechner specification popularized by Hey and Orme [1994] implies a different but equally simple change to `ML_eut0`. Again we add an error term “mu” to the arguments of the program, as above, and now we have the latent index

```
generate double `euDiff' = (`euR' - `euL')/`mu'
```

instead of the original

```
generate double `euDiff' = `euR' - `euL'
```

Here are the results:

```

. ml model lf ML_eut (r: Choices P0left P1left P2left P3left P0right P1right P2right
P3right prize0 prize1 prize2 prize3 stake = ) (mu: ), cluster(id) maximize

. ml display
                                Number of obs   =          9311
                                Wald chi2(0)      =           .
                                Prob > chi2       =           .

                                (Std. Err. adjusted for 158 clusters in id)
-----+-----
                                |               |               |               |               |               |
                                |           Coef.   |   Robust      |           z     |           P>|z|   |           [95% Conf. Interval]
                                |-----+-----|-----+-----|-----+-----|-----+-----|
r                                |           _cons   |           .7709795 |           .0439284 |           17.55   |           0.000   |           .6848814   |           .8570776
mu                                |           _cons   |           .9660444 |           .1336743 |           7.23    |           0.000   |           .7040477   |           1.228041
-----+-----
. ml display

```

So the CRRA coefficient declines very slight, and the noise term is estimated as a normal probability with standard deviation of 0.966.

B4. Extensions

There are many possible extensions of the basic programming elements considered here.

Harrison [2006] illustrates the following:

- modeling non-parametric utility functions, following Hey and Orme [1994];
- modeling rank-dependent decision weights for the RDU and RDEV structural model;
- modeling rank-dependent decision weights and sign-dependent utility for the CPT structural model;
- the imposition of constraints on parameters to ensure non-negativity (e.g., $\lambda > 1$ or $\mu > 0$), or finite bounds (e.g., $0 < r < 1$);
- the specification of finite mixture models;
- the coding of non-nested hypothesis tests; and
- maximum simulated likelihood, in which one or more parameters are treated as random coefficients to reflect unobserved individual heterogeneity (e.g., Train [2003]).

In each case template code is provided along with data and illustrative estimates.

Additional References

- Gould, William; Pitblado, Jeffrey, and Sribney, William, *Maximum Likelihood Estimation With Stata* (College Station, TX: Stata Press, Third Edition, 2006).
- Harrison, Glenn W., "Maximum Likelihood Estimation of Utility Functions Using *Stata*," *Working Paper 06-12*, Department of Economics, College of Business Administration, University of Central Florida, 2006.
- Rabe-Hesketh, Sophia, and Everitt, Brian, *A Handbook of Statistical Analyses Using Stata* (New York: Chapman & Hall/CRC, Third Edition, 2004).
- Paarsch, Harry J., and Hong, Han, *An Introduction to the Structural Econometrics of Auction Data* (Cambridge, MA: MIT Press, 2006).
- Train, Kenneth E., *Discrete Choice Methods with Simulation* (New York: Cambridge University Press, 2003).

Appendix C: Instructions for Laboratory Experiments

(NOT FOR PUBLICATION)

C1. Baseline Instructions

YOUR INSTRUCTIONS

This is an experiment that is just like the television program *Deal Or No Deal*. There are 26 prizes in 26 suitcases. You will start by picking one suitcase, which becomes “your case.” Then you will be asked to open some of the remaining 25 cases. At the end of each round, you will receive a “bank offer” to end the game. If you accept the bank offer, you will receive that money. If you turn down the bank offer, you will be asked to open a few more cases, and then there will be another bank offer. The bank offer depends on the value of the prizes that have not been opened. If you pick cases that have higher prizes, the next bank offer will tend to be lower; but if you pick cases that have lower prizes, the next bank offer will tend to be higher.

Lets go through an example of the game. All of the screen shots here were taken from the game. The choices were ones we made just to illustrate. Your computer will provide better screen shots that are easier to read. The *Deal Or No Deal* logo shown on these screens is the property of the production company of the TV show, and we are not challenging their copyright.

Here is the opening page, telling you how many cases you have to open in each round. The information on your screen may differ from this display, so be sure to read it before playing the game. Click on Begin to start the game...

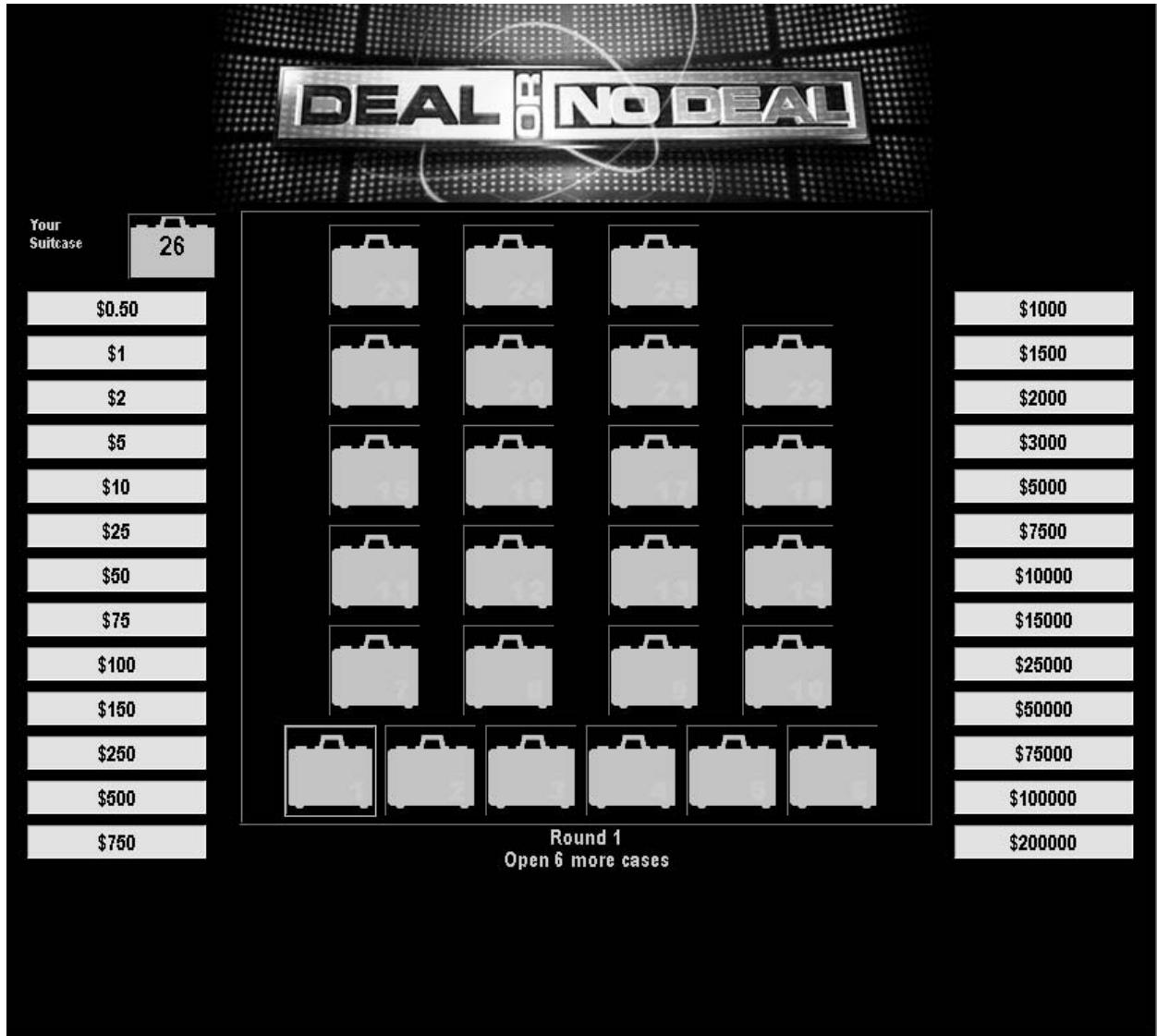


Here is what the opening page looks like. There are 26 cases displayed in the middle. There are also 26 cash prizes displayed on either side. Each case contains one of these prizes, and each prize is in one of the cases. We will convert your earnings in this game into cash at an exchange rate of 1000 to 1. So if you earn \$200,000 in the game you will receive \$200 in cash, and if you earn \$1,000 in the game you will receive \$1 in cash. We will round your earnings to the nearest penny when we use the exchange rate.



The instruction at the bottom asks you to choose one of the suitcases. You do this by clicking on the suitcase you want.

Here we picked suitcase 26, and it has been moved to the top left corner. This case contains one of the prizes shown on the left or the right of the screen.



Round 1 of the game has now begun. You are asked to open 6 more cases. You do this by clicking on the suitcases you want to open.

In this round we picked suitcases 2, 4, 10, 15, 17 and 24. It does not matter which order they were picked in. You can see the prizes in these cases revealed to you as you open them. They are revealed and then that prize on the left or right is shaded, so you know that this prize is not in your suitcase.



At the end of each round the bank will make you an offer, listed at the bottom of the screen. In this case the bank offers \$4,000. If you accept this deal from the bank, by clicking on the green DEAL box, you receive \$4,000 and the game ends. If you decide not to accept the offer, by clicking on the red NO DEAL box, you will go into round 2.

We will wait until everyone has completed their game before continuing today, so there is no need for you to hurry. If you choose DEAL we ask that you wait quietly for the others to finish.

To illustrate, we decided to say NO DEAL in round 1 and picked 5 more cases. You can see that there are now more shaded prizes, so we have a better idea what prize our suitcase contains. The bank made another offer, in this case \$4,900.

DEAL OR NO DEAL

Your Suitcase: 26

Prize List (Left): \$0.50, \$1, \$2, \$5, \$10, \$25, \$50, \$75, \$100, \$150, \$250, \$500, \$750

Prize List (Right): \$1000, \$1500, \$2000, \$3000, \$5000, \$7500, \$10000, \$15000, \$25000, \$50000, \$75000, \$100000, \$200000

Suitcase Values: 23, 20, 21, 16, 18, 11, 12, 13, 14, 8, 9, 1, 3, 5

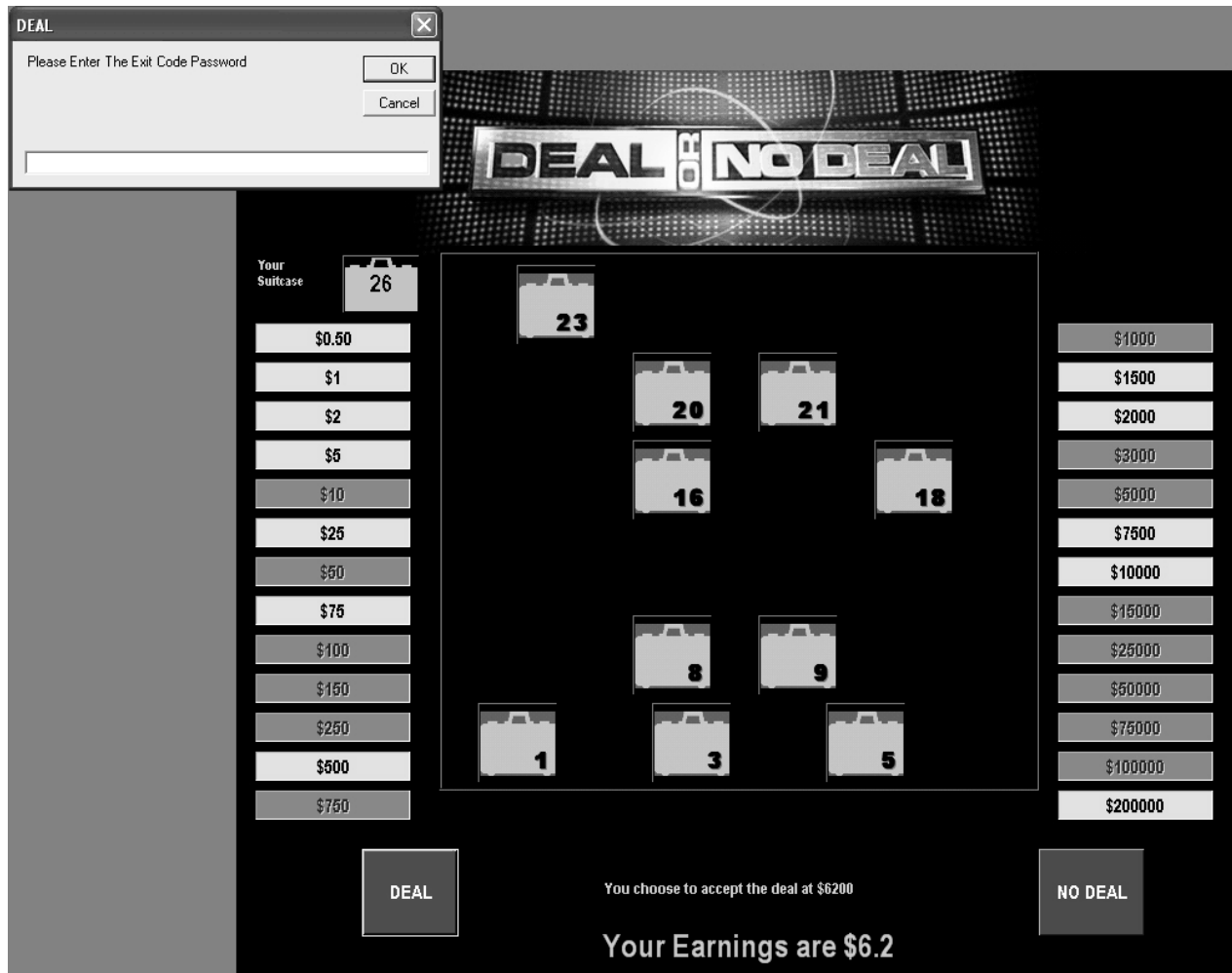
You have opened 11 suitcases
 There are 15 unopened suitcases left
 If you do not take this offer you have to open 4 more cases before the bank gives you another offer

The bank offers you \$4900

Buttons: DEAL, NO DEAL

If we accept this bank offer the screen tells us our cash earnings. We went on and played another round, and then accepted a bank offer of \$6,200. So our cash earnings are $\$6,200 \div 1,000 = \6.20 , as stated on the screen.

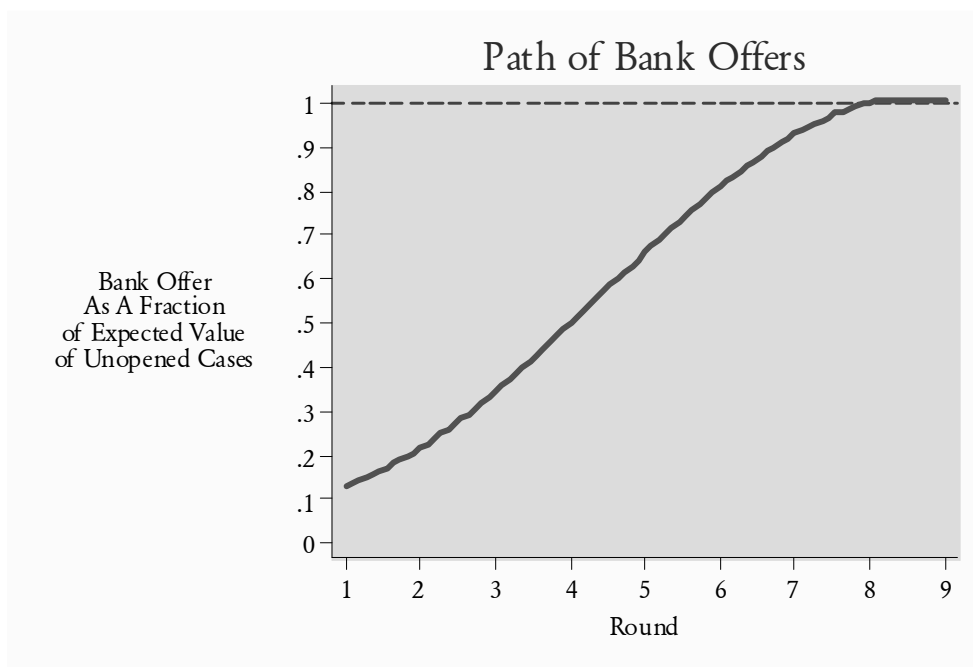
A small box then appears in the top left corner. This is for the experimenter to fill in. Please just signal the experimenter when you are at this stage. They need to write down your cash earnings on your payoff sheet, and then enter the super-secret exit code. Please do not enter any code in this box, or click OK.



The bank's offers depend on two things. First, how many rounds you have completed. Second, the expected value of the prizes in the unopened cases.

The expected value is simply the average earnings you would receive at that point if you could open one of the unopened prizes and keep it, and then go back to the original unopened prizes and repeat that process many, many times. For example, in the final round, if you have prizes of \$1,000 and \$25,000 left unopened, your expected value would be $(\$1,000 \times \frac{1}{2}) + (\$25,000 \times \frac{1}{2}) = \$26,000 \div 2 = \$13,000$. You would get \$1,000 roughly 50% of the time, and you would get \$25,000 roughly 50% of the time, so the average would be \$13,000. It gets a bit harder to calculate the expected value with more than two unopened prizes, but the idea is the same.

As you get to later rounds, the bank's offer is more generous in terms of the fraction of this expected value. The picture below shows you the path of bank offers. Of course, your expected value may be high or low, depending on which prizes you have opened. So 90% of a low expected value will generate a low bank offer in dollars, but 50% of a high expected value will generate a high bank offer in dollars.



There is no right or wrong choice. Which choices you make depends on your personal preferences. The people next to you will have different lotteries, and may have different preferences, so their responses should not matter to you. Nor do their choices affect your earnings in any way. Please work silently, and make your choices by thinking carefully about the bank offers.

All payoffs are in cash, and are in addition to the \$5 show-up fee that you receive just for being here. Do you have any questions?

C2. Additional Instructions for UK Version

The version of the game you will play now has several differences from the practice version.

First, there will be 7 rounds in the version instead of the 10 rounds of the practice. This is listed on the screen displayed below.

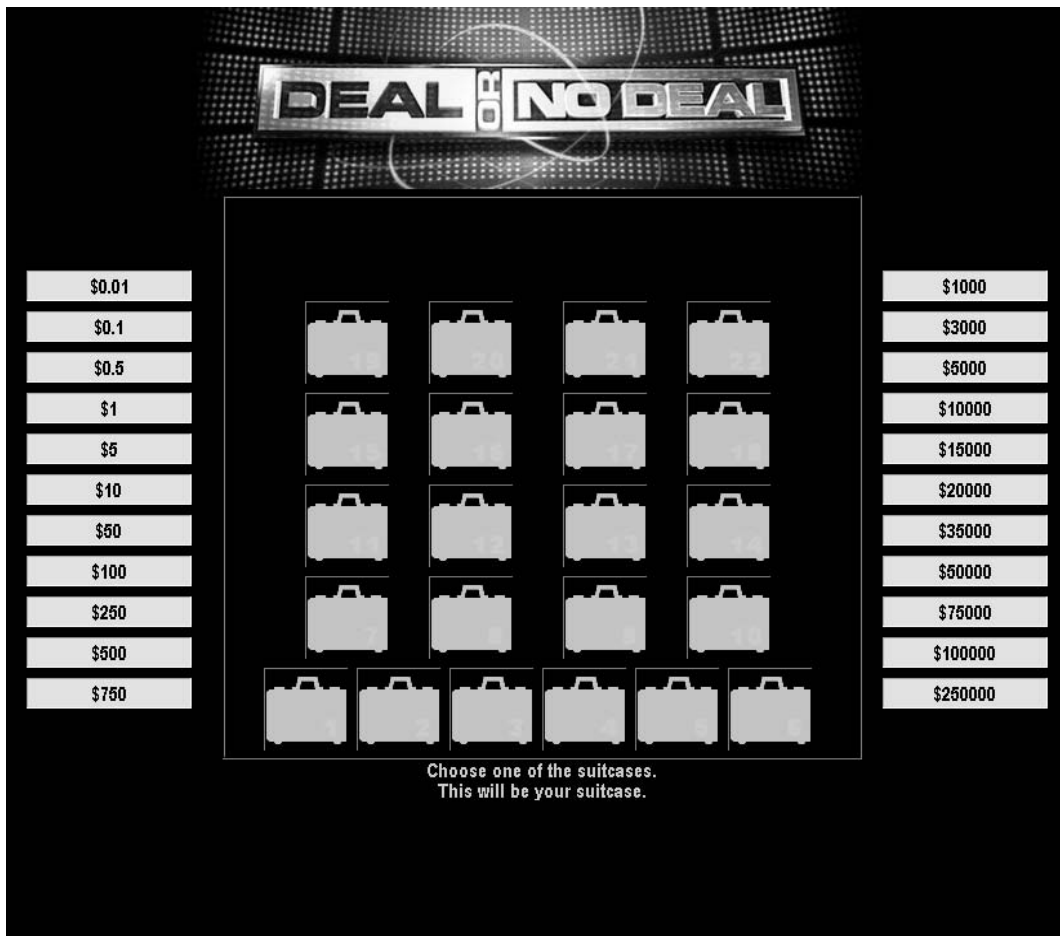
Second, you will open suitcases in the sequence shown in the screen displayed below.

Third, there are only 22 prizes instead of 26 in the practice, and the prize amounts differ from the prizes used in the practice. The prizes are listed on the screen displayed on the next page.

Fourth, the bank's offer function is the one displayed on the next page. It still depends on how many rounds you have completed.

In all other respects the game is the same as the practice. Do you have any questions?





Path of Bank Offers

Bank Offer
As A Fraction
of Expected prize
of Unopened Cases

