MPRA
Munich Personal RePEc Archive

Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk

Wilcox, Nathaniel
University of Houston

November 2007

‘Stochastically More Risk Averse:’ A Contextual Theory of Stochastic Discrete Choice Under Risk

by

Nathaniel T. Wilcox*
University of Houston

Forthcoming in the Journal of Econometrics

Abstract

Microeconometric treatments of discrete choice under risk are typically homoscedastic latent variable models. Specifically, choice probabilities are given by preference functional differences (given by expected utility, rank-dependent utility, etc.) embedded in cumulative distribution functions. This approach has a problem: Estimated utility function parameters meant to represent agents’ degree of risk aversion in the sense of Pratt (1964) do not imply a suggested “stochastically more risk averse” relation within such models. A new heteroscedastic model called “contextual utility” remedies this, and estimates in one data set suggest it explains (and especially predicts) as well or better than other stochastic models.

JEL Classification Codes: C25, C91, D81

Keywords: risk, more risk averse, discrete choice, stochastic choice, heteroscedasticity.

*Department of Economics, University of Houston, Houston, TX 77204-5019, phone 713-743-3840, fax 713-743-3798, email nwilcox@mail.uh.edu.

There is no doubt of the importance of two papers in the development of expected utility or EU theory. Pratt (1964) gave us an EU-based understanding of "agent *a* is more risk averse than agent *b*:" Write this as $a \underset{mra}{\succ} b$ and call this relation "MRA in Pratt's sense." Pratt also gave us risk aversion measures and parametric utility functions for money outcomes that represent these measures of risk aversion in their parameters. Rothschild and Stiglitz (1970) then considered possible definitions of the relation "lottery *T* is riskier than lottery *S*" and proved that several definitions are equivalent under EU. Call $\{S,T\}$ an MPS pair when *T* is a mean preserving spread of *S* (defined subsequently). Rothschild and Stiglitz showed that any EU agent with a strictly concave utility of money prefers *S* to *T* when choosing from any MPS pair $\{S,T\}$.

An accumulated experimental literature suggests various problems with EU, and this has spawned alternative structural theories such as prospect theory (Kahneman and Tversky 1979), rank-dependent utility or RDU (Quiggin 1982; Chew 1983) and cumulative prospect theory or CPT (Tversky and Kahneman 1992). However, the experimental literature also established a fact that is not directly addressed by these theories: Choice under risk appears to be highly stochastic. Beginning with Mosteller and Nogee (1951), experiments with repeated trials of pairs reveal substantial choice switching by the same subject between trials. In some cases, the trials span days (e.g. Tversky 1969; Hey and Orme 1994; Hey 2001) and one might worry that decision-relevant conditions may have changed between trials. Yet substantial switching occurs even between trials separated by bare minutes, with no intervening change in wealth, background risk, or any other obviously decision-relevant variable (Camerer 1989; Starmer and Sugden 1989; Ballinger and Wilcox 1997; Loomes and Sugden 1998).

How do we generalize the relation "more risk-averse" to stochastic choice under risk? Suppose P^n is the probability that agent *n* chooses *S* from MPS pair $\{S,T\}$, and suppose we regard these probabilities as the theoretical primitive. What might it mean for agent *a* to be stochastically more risk-averse than agent *b*? Consider this proposal:

Stochastically More Risk-Averse (SMRA). Agent *a* is stochastically more risk averse than agent *b*, written $a \underset{smra}{\succ} b$, *iff* $P^a > P^b$ for every MPS pair $\{S,T\}$.[1]

That is, the stochastically more risk averse agent is more likely to choose the relatively safe lottery in every MPS pair. If sample proportions of relatively safe choices from MPS pairs vary significantly across subjects in an experiment, and we call this heterogeneous risk aversion, then we probably have something like SMRA in mind.

Today we are witnessing rapid growth of a "new structural econometrics" of discrete choice under risk, usually based on discrete choices from lottery pairs. In practice, a parametric functional form from Pratt (1964) or a generalization of them (Saha 1993) is used to specify an EU, RDU or CPT value difference between lotteries in a pair—what I will call a *V*-difference. These *V*-differences are then embedded within a c.d.f. to specify choice probabilities for maximum likelihood or other *M*-estimation. Call this approach a *V*-difference latent variable model. I argue that there is a deep problem with this approach: Parameters in such models that are meant to represent degrees of risk aversion in Pratt's sense cannot represent degrees of stochastic risk aversion across all agents and all decision contexts in the SMRA sense.

For example, suppose we use a *V*-difference latent variable model to estimate coefficients of relative risk aversion ρ for Anne and Bob, getting estimates $\hat{\rho}^{Anne} > \hat{\rho}^{Bob}$. We would like to say that Anne $\underset{mra}{\succ}$ Bob in Pratt's sense. However, sections 3 and 4 below will show that the *V*-

---

[1] To my knowledge, Hilton (1989) first suggested definitions of "more risk-averse" for stochastic choice. Let $EV(T)$ be the expected value of lottery *T* and consider pairs of the form $\{EV(T),T\}$. Hilton defined "agent *a* is more risk-averse in selection than agent *b*" as $P^a > P^b$ for all pairs $\{EV(T),T\}$. SMRA simply generalizes Hilton's definition from a specific kind of MPS pair (a choice between a lottery and its expected value) to any MPS pair.

difference latent variable model and $\rho^{Anne} > \rho^{Bob}$ cannot imply that Anne $\underset{smra}{\succ}$ Bob in the sense defined above. This occurs because the utility or value of money $u(z)$ in theories such as EU, RDU and CPT is only unique up to an affine transformation and, therefore, it is not a theorem that MRA in Pratt's sense implies a greater *V*-difference between lotteries in all MPS pairs.

However, MRA in Pratt's sense does imply orderings of ratios of differences of utilities. Psychologically, when we say Anne $\underset{mra}{\succ}$ Bob in Pratt's sense, on some interval of outcomes $[z_1, z_3]$, we mean that Anne perceives the ratio $[u(z_2)-u(z_1)]/[u(z_3)-u(z_1)]$ as larger than Bob does for all $z_2 \in (z_1, z_3)$. Economically, this is an implication of Pratt's main theorem. Therefore, both economics and psychology suggest that if we wish MRA in Pratt's sense to imply SMRA in a latent variable model, the latent variable will of necessity involve ratios of utility differences rather than utility differences. Put differently, for MRA in Pratt's sense to imply SMRA, our stochastic model may need to assume that agents perceive EU, RDU or CPT *V*-differences relative to some salient utility difference. In section 5, I suggest a contextual utility stochastic choice model: It assumes that this salient utility difference is the difference between the utilities of the maximum and minimum possible outcomes in a lottery pair. This model will allow SMRA to be an implication of MRA in Pratt's sense, in suitably defined ways. More generally, it seems that when we move from deterministic to stochastic choice under risk, context and risk aversion are inextricably entwined with one another. Put differently: If choice under risk is stochastic, a globally coherent notion of greater risk aversion necessarily implies that context matters.

As a relational preference concept, MRA is ubiquitous in professional economic discourse. Figure 1 illustrates this for the years 1977 to 2001, comparing counts of articles with either of the phrases "more risk-averse" or "greater risk aversion" in their text to counts of articles with either of the phrases "more substitutable" or "greater elasticity of substitution" in their text.[2] The latter is clearly a central relational concept of economics. The comparative levels of the two trend lines are somewhat arbitrary, since searches for different text strings would affect the levels. Moreover, instances of the substitutability phrases include articles about technologies as well as preferences. The comparative trends are more meaningful. MRA gained on "more substitutable" as a relational concept over the quarter century from 1977 to 2001. Figure 1 also shows that although citations[3] to Pratt (1964) peaked in the early 1980s, they have continued at a fairly constant rate of about thirty-five to forty a year over the last two decades and show no sign of decrease. Probably, very few economics articles show such steady and long-lived influence.

Given the ubiquity of MRA in economic discourse, it would be nice if model parameters meant to represent MRA in Pratt's sense had a theoretically appealing observable meaning such as SMRA. Contextual utility delivers such meaning: Common alternative models do not. Section 6 shows that contextual utility may be empirically better too: Both EU and RDU models estimated with the contextual utility stochastic model explain (and especially predict) risky choices better than *V*-difference latent variable models and the random preference stochastic model in the well-known Hey and Orme (1994) data set.

1. Preliminaries

Let the vector $c = (z_{1c}, z_{2c}, z_{3c})$ be three money outcomes with $z_{1c} < z_{2c} < z_{3c}$, called an <u>outcome context</u> or simply a <u>context</u>. Let S_m be a discrete probability distribution (s_{m1}, s_{m2}, s_{m3})

[2] Some journals withhold JSTOR availability for recent years: 2001 is the most recent widely available year. Some important journals either do not have JSTOR availability, or did not exist, back to 1977 (e.g. the <u>Journal of Labor Economics</u> and the <u>Journal of Economic Perspectives</u>, respectively) and so are not included. Three important business journals (the <u>Journal of Business</u>, <u>Journal of Finance</u> and <u>Management Science</u>) are included as well.

[3] The citation counts are from 1977 to 2006 in all fields (not just economics) in the ISI Science and Social Science Citation indices. The temporal pattern of citations in economics publications alone is essentially similar.

on a vector (z_1, z_2, z_3); then a three-outcome lottery S_{mc} on context *c* is the distribution S_m on the outcome vector (z_{1c}, z_{2c}, z_{3c}). Sometimes lotteries are viewed as cumulative distribution functions $S_{mc}(z) \equiv \sum_{i|z_{ic} \le z} s_{mi}$. A pair *mc* is two distinct lotteries $\{S_{mc}, T_{mc}\}$ on context *c*, and a basic pair is one where neither lottery first-order stochastically dominates the other. In basic pairs, we may choose lottery names so that $s_{m2} + s_{m3} > t_{m2} + t_{m3}$ and $s_{m3} < t_{m3}$, and say that S_{mc} is safer than T_{mc} in the sense that S_{mc} has more chance of the center outcome z_{2c}, and less chance of the extreme outcomes z_{1c} and z_{3c}, than does T_{mc}. If the expected values $E(S_{mc})$ and $E(T_{mc})$ are equal, the pair is a mean-preserving spread or MPS pair. Let Ω_{mps} be the set of all MPS pairs: Risk-averse EU agents prefer S_{mc} to T_{mc} $\forall\, mc \in \Omega_{mps}$ (Rothschild and Stiglitz 1970).

To connect deterministic theory to probabilistic primitives, let the structure of choice under risk be defined as a function *V* of lotteries and a vector of structural parameters β^n such that

$$V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n) \ge 0 \Leftrightarrow P^n_{mc} \ge 0.5, \quad (1.1)$$

[4]

where P^n_{mc} is the probability that agent *n* chooses S_{mc} from pair *mc*. Call $V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n)$ the *V*-difference in pair *mc*: When this is nonnegative, it represents the deterministic primitive "*n* weakly prefers S_{mc} to T_{mc}." In turn, (1.1) equates this with the probabilistic primitive "*n* is not more likely to choose T_{mc} from *mc*," a common approach since Edwards (1954).[5]

For example, expected utility (EU) with the constant relative risk aversion (CRRA) utility of money $u^n(z) = (1-\rho^n)^{-1} z^{1-\rho^n}$ is the structure $V(S_{mc} \mid \rho^n) = (1-\rho^n)^{-1} \sum_{i=1}^{3} s_{mi} z_{ic}^{1-\rho^n}$, so that

$$(1-\rho^n)^{-1} \sum_{i=1}^{3} (s_{mi} - t_{mi}) z_{ic}^{1-\rho^n} \ge 0 \Leftrightarrow P^n_{mc} \ge 0.5, \quad (1.2)$$

[4] This restricts attention to transitive structures; this would be an unsatisfactory equation for a nontransitive theory.
[5] There are stochastic choice models under which this innocent-sounding equation is incorrect, such as Machina (1985). Though as yet scant, existing evidence is not promising for these alternatives (Hey and Carbone 1995).

where $\beta^n = \rho^n$ is the sole structural parameter, called agent *n*'s coefficient of relative risk aversion. Call this the CRRA EU structure. The rank-dependent utility (or RDU) structure (Quiggin 1982; Chew 1983) replaces the probabilities s_{mi} in EU with weights ws_{mi}. These weights are $ws_{mi} = w\left(\sum_{z\geq i} s_{mz}\right) - w\left(\sum_{z>i} s_{mz}\right)$, where a continuous and strictly increasing weighting function $w(q)$ takes the unit interval onto itself. Writers suggest several parametric forms for the weighting function; here, I use Prelec's (1998) one-parameter form, which is $w(q \mid \gamma^n) = \exp(-[-\ln(q)]^{\gamma^n})\ \forall\ q \in (0,1)$, $w(0)=0$ and $w(1)=1$. Then a CRRA RDU structure is $V(S_{mc} \mid \rho^n, \gamma^n) = (1-\rho^n)^{-1}\sum_{i=1}^{3} ws_{mi}(\gamma^n) z_{ic}^{1-\rho^n}$, so that

$$(1-\rho^n)^{-1}\sum_{i=1}^{3}[ws_{mi}(\gamma^n) - wt_{mi}(\gamma^n)] z_{ic}^{1-\rho^n} \geq 0 \Leftrightarrow \mathrm{P}_{mc}^n \geq 0.5 . \qquad (1.3)$$

Structures with weighting functions attribute risk attitudes to both utility function and weighting function shape (Quiggin 1982; Tversky and Kahneman 1992). This is why I have (so far) written "MRA in Pratt's sense:" My issue is the partial contribution of utility function shape to risk attitude, holding weighting functions constant. Henceforth, when I write $a \underset{mra}{\succ} b$ and $a \underset{smra}{\succ} b$, or refer to the MRA or SMRA relation, I am always considering agents with identical weighting functions. The constant reminder "in Pratt's sense" will now cease. However, a certain fact about RDU weights in basic pairs will be used in section 5 to derive properties of the contextual utility model. Since $w(q)$ is strictly increasing, and since $s_{m2} + s_{m3} > t_{m2} + t_{m3}$ and $s_{m3} < t_{m3}$ in basic pairs, $w(s_{m2} + s_{m3}) > w(t_{m2} + t_{m3})$ and $w(s_{m3}) < w(t_{m3})$. So in all basic pairs,

$$ws_{m2} - wt_{m2} \equiv w(s_{m2} + s_{m3}) - w(s_{m3}) - [w(t_{m2} + t_{m3}) - w(t_{m3})] > 0 . \qquad (1.4)$$

2. *V*-difference latent variable models: Strong and strict utility

To estimate any vector β^n in (1.1), we need a stochastic model to complete the relationship between *V*-difference and choice probability. Latent variable models are one way to do this. Let $y^n_{mc} = 1$ if agent *n* chooses S_{mc} from pair *mc* ($y^n_{mc} = 0$ otherwise). In general, such models assume that there is an underlying but unobserved continuous random latent variable y^{n*}_{mc} such that $y^n_{mc} = 1 \Leftrightarrow y^{n*}_{mc} \geq 0$; then we have $P^n_{mc} = \Pr(y^{n*}_{mc} \geq 0)$. Here, the latent variable is

$$y^{n*}_{mc} = V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n) - \varepsilon / \lambda, \tag{2.1}$$

where ε is a mean zero random variable with some standard variance and c.d.f. $H(x)$ such that $H(0) = 0.5$ and $H(x) = 1 - H(-x)$, usually assumed to be the standard normal or logistic c.d.f. The resulting latent variable model of P^n_{mc} is then

$$P^n_{mc} = H\big(\lambda[V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n)]\big). \tag{2.2}$$

In decision theory, (2.2) is a strong utility model (Debreu 1958; Block and Marschak 1960; Luce and Suppes 1965). Letting *L* be the set of all lotteries, choice probabilities follow strong utility if there is a function $\mu^n : L \rightarrow \mathbb{R}$ and an increasing function $\varphi^n : \mathbb{R} \rightarrow [0,1]$, with $\varphi^n(0) = 0.5$ and $\varphi^n(x) = 1 - \varphi^n(-x)$, such that $P^n_{mc} = \varphi^n\big(\mu^n(S_{mc}) - \mu^n(T_{mc})\big)$. Clearly, (2.2) is a strong utility model where $\varphi^n(x) = H(\lambda x)$ and $\mu^n(S_{mc}) = V(S_{mc} \mid \beta^n)$. In (2.1), ε / λ may be regarded as computational, perceptual or evaluative noise in the decision maker's apprehension of the *V*-difference $V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n)$, with λ^{-1} proportional to the standard deviation of this noise. Microeconometric doctrine usually views ε / λ as a perturbation to *V*-difference observed by agents but not observed by the econometrician. In either case, as λ approaches infinity, choice probabilities converge on either zero or one, depending on the sign of the *V*-difference; put differently, the observed choice becomes increasingly likely to express the underlying preference direction of the structure. We can call λ^{-1} the noise parameter or call λ the precision parameter.

Strict utility (Luce and Suppes 1965) is more restrictive. Again there is a scale μ^n defined on lotteries, but it must be strictly positive and choice probabilities must take the form

$$P^n_{mc} = \mu^n(S_{mc})/[\mu^n(S_{mc}) + \mu^n(T_{mc})]. \tag{2.3}$$

As Luce and Suppes point out (p. 335), every strict utility model is algebraically identical to

$$P^n_{mc} = \Lambda\big(\ln[\mu^n(S_{mc})] - \ln[\mu^n(T_{mc})]\big), \text{ where } \Lambda(x) = (1+e^{-x})^{-1} \text{ is the logistic c.d.f.} \tag{2.4}$$

If V is positive-valued, we could for instance choose $\mu^n(S_{mc}) \equiv V(S_{mc} \mid \beta^n)^\lambda$ and rewrite (2.4) as

$$P^n_{mc} = \Lambda\Big(\lambda\big(\ln[V(S_{mc} \mid \beta^n)] - \ln[V(T_{mc} \mid \beta^n)]\big)\Big) \tag{2.5}$$

This resembles (2.2) except that a difference of logarithms of V, or logarithmic V-difference, replaces the V-difference. Model (2.5) was employed by Holt and Laury (2002). Another very common strict utility form, used both for quantal response equilibrium (McKelvey and Palfrey 1995) and the experience-weighted attraction learning model (Camerer and Ho 1999), sets $\mu^n(S_{mc}) \equiv \exp[\lambda V(S_{mc} \mid \beta^n)]$. Substituting this into (2.4) gives

$$P^n_{mc} = \Lambda\big(\lambda[V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n)]\big), \tag{2.6}$$

which is identical to the strong utility model (2.2) with the logistic c.d.f. as the choice of H. My strong utility estimations will in fact be (2.2) with the logistic c.d.f., and hence also equivalent to this very common strict utility form. What I will call my strict utility estimations instead use the logarithmic V-difference form and logistic c.d.f., as in (2.5) and Holt and Laury.

Homoscedasticity with respect to pairs mc (constant λ across pairs and contexts) is the essence of strong and strict utility models: Without it, they do not imply strong stochastic transitivity or SST, their definitive property (Block and Marschak 1960; Luce and Suppes 1965).

However, homoscedasticity with respect to agents n is not, and various transformations τ of V-difference may be interpreted as agent heteroscedasticity determined by risk parameters. For example, call the textbook CRRA utility function $u_b(z \mid \rho) = (1-\rho)^{-1} z^{1-\rho}$ the basic transformation, and call $u_p(z \mid \rho) = z^{1-\rho}$ the power transformation. When we write the V-difference in a strong utility model using the basic transformation, this is algebraically identical to writing the V-difference with the power transformation and additionally assuming agent-specific precision parameters λ^n that are proportional to $|(1-\rho^n)|$. Section 4 will take up agent heteroscedasticity in a more theoretically grounded manner.

In experiments with $K \geq 4$ distinct outcomes $z_1 < z_2 < ... < z_K$ (and several distinct three-outcome contexts), call $u_{ur}(z \mid \rho) = (z^{1-\rho} - z_1^{1-\rho})/(z_K^{1-\rho} - z_1^{1-\rho})$ the unit range transformation (making the utility range across all outcomes equal to unity for all ρ) and call $u_{mu}(z \mid \rho) =$ $(z^{1-\rho} - z_1^{1-\rho})/(z_2^{1-\rho} - z_1^{1-\rho})$ the minimum unit transformation (making the utility difference between the two smallest outcomes equal to unity for all ρ). With minor terminological abuse, I will say that the logarithmic V-difference latent variable in (2.5) is written using the logarithmic transformation. Let $\Delta_\tau V_{mc}(\rho)$ be the CRRA EU V-difference (or logarithmic V-difference) in pair mc, as written with any of these five transformations τ. With any constant λ, $\Delta_\tau V_{mc}(\rho)$ would need to be monotone increasing in ρ for strong or strict utility models to imply SMRA.

3. Critique of strong and strict utility

Pratt's (1964) original EU definition of $a \underset{mra}{\succ} b$ was extended by Chew, Karni and Safra (1987, p. 374) for RDU. I quote them here at length, with notation modified suitably to fit mine:

To compare the attitudes toward risk of two preference relations $\succeq^b$ and $\succeq^a$ on D_J [a set of probability distributions on an interval $J \subset \mathbb{R}$], we define $T_{mc} \in D_J$ to differ from $S_{mc} \in D_J$ by a simple compensated spread from the viewpoint of $\succeq^b$, if and only if $T_{mc} \sim^b S_{mc}$ and $\exists\, z^0 \in J$ such that $T_{mc}(z) \geq S_{mc}(z)$ for all $z < z^0$ and $T_{mc}(z) \leq S_{mc}(z)$ for all $z \geq z^0$.

DEFINITION 5: A preference relation $\succeq^a \in \mathbb{G}$ [the set of preference relations representable by Gateaux-differentiable RDU functionals V] is said to be *more risk averse than* $\succeq^b \in \mathbb{G}$ if $S_{mc} \succeq^a T_{mc}$ for every $T_{mc}, S_{mc} \in D_J$ such that T_{mc} differs from S_{mc} by a simple compensated spread from the point of view of $\succeq^b$….

THEOREM 1. *The following conditions on a pair of Gateaux differentiable* [RDU] *preference functionals V^b and V^a on D_J with respective utility functions u^b and u^a and* [*weighting*] *functions w^b and w^a are equivalent*:…

(ii) *w^a and u^a are concave transformations of w^b and u^b, respectively.*

(iii) *If T_{mc} differs from S_{mc} by a simple compensated spread from the point of view of $\succeq^b$* [implying that $V^b(T_{mc}) = V^b(S_{mc})$] *then* $V^a(T_{mc}) \leq V^a(S_{mc})$ [all italics in original].

Consider what this means for a CRRA EU V-difference. In this case both of the weighting functions w^a and w^b are identity functions; and in every MPS pair, T_{mc} will be a simple compensated spread of S_{mc} from the viewpoint of a risk-neutral decision maker, that is an agent b with $\rho^b = 0$, so that $V^b(T_{mc}) = V^b(S_{mc})$. Equivalently, we may write $\Delta_\tau V_{mc}(0) = 0$ in any MPS pair. The results (ii) and (iii) of Theorem 1 then let us conclude that $V^a(T_{mc}) \leq V^a(S_{mc})$ for any more risk-averse agent a, in this case any agent a with $\rho^a > 0$. Equivalently, we may write $\Delta_\tau V_{mc}(\rho^a) \geq 0\ \forall\ \rho^a > 0$ in any MPS pair. However, nothing in the theorem implies that $\partial \Delta_\tau V_{mc}(\rho)/\partial\rho > 0\ \forall\ \rho > 0$: Nothing suggests that any V-difference, written using any transformation, is monotone increasing in any parameter meant to represent MRA.

The theorem only says that indifference becomes weak preference when we substitute any "more risk averse" utility function (or probability weighting function) for the less risk averse one that generated indifference: It says nothing about monotone increasing *V*-difference with greater risk aversion. But the latter is exactly what is required for $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$ in strong and strict utility models when λ is constant across agents. In particular, strong utility specifies no necessary relationship between ρ and λ, so CRRA EU strong utility and MRA do not imply SMRA.

Moreover, it seems that MRA <u>cannot</u> imply SMRA in <u>any</u> strong or strict utility model that assumes constant λ across agents. Graphs illustrate this using MPS pairs from Hey's (2001) four-outcome design (these are common, e.g. Hey and Orme 1994 and Harrison and Rutström 2005). The design employs the four outcomes 0, 50, 100 and 150 U.K. pounds and these generate four distinct three-outcome contexts: All pairs are on one of these contexts. Index contexts by their omitted outcome, e.g. context $c = -0$ is (50,100,150), while context $c = -150$ is (0,50,100).

Figure 2 shows how $\lambda_\tau \Delta_\tau V_{1,-150}(\rho)$ behaves for the five transformations τ defined previously,[6] computed for the MPS pair 1 on context $c = -150$, given by $S_{1,-150} = (0,7/8,1/8)$ and $T_{1,-150} = (3/8,1/8,4/8)$ taken from Hey (2001), as ρ varies over the interval [0,0.99]. To sketch Figure 2, a constant value of λ_τ has been chosen for <u>each</u> transformation to make $\lambda_\tau \Delta_\tau V_{1,-150}(\rho)$ reach a common maximum of 10 for $\rho \in$ [0,0.99], for easy visual comparisons; the important point is that λ_τ is held constant to draw each graph of $\lambda_\tau \Delta_\tau V_{1,-150}(\rho)$, for each transformation τ. If $\lambda_\tau \Delta_\tau V_{1,-150}(\rho)$ is nonmonotone on [0,0.99], then MRA <u>cannot</u> imply SMRA under transformation τ given constant λ across agents.

---

[6] Here, the CRRA unit range transformation is $u_{ur}(z \mid \rho) = (z^{1-\rho} - 0^{1-\rho})/(150^{1-\rho} - 0^{1-\rho}) = (z/150)^{1-\rho}$, while the minimum unit transformation is $u_{mu}(z \mid \rho) = (z^{1-\rho} - 0^{1-\rho})/(50^{1-\rho} - 0^{1-\rho}) = (z/50)^{1-\rho}$.

Figure 2 shows that the basic, power and log transformations are not monotonically increasing in ρ, though the nonmonotonicity for the basic transformation is mild for this particular pair. The implication is that in a strong utility CRRA EU model with λ constant across agents, MRA defined in terms of ρ cannot imply SMRA using either the basic, power or log transformations, for MPS pair 1 on context $c = -150$. However, the unit range and minimum unit transformations can do this, at least for this MPS pair on this context.

Unfortunately, this property of the unit range and minimum unit transformations disappears as soon as we consider a context that does not share the minimum outcome $z_1 = 0$ used to define those transformations. Consider now MPS pair 1 on context $c = -0$, that is (50,100,150), given by $S_{1,-0} = (0,7/8,1/8)$ and $T_{1,-0} = (3/8,1/8,4/8)$, also taken from Hey (2001). Figure 3 shows that for this MPS pair, $\lambda_\tau \Delta_\tau V_{1,-0}(\rho)$ is severely nonmonotonic for all five transformations. MRA defined in terms of ρ cannot imply SMRA with any of these five transformations, given constant λ across agents, across all contexts constructed from a four-outcome vector. If we want to explain agent heterogeneity of safe choices from MPS pairs by means of a strong or strict utility model solely in terms of a risk parameter like ρ, and we use any of the five transformations discussed above, we cannot succeed if we assume that precision λ is constant across agents. It seems that $a \underset{mra}{\succ} b \not\Rightarrow a \underset{smra}{\succ} b$ in strong and strict utility models with λ constant across agents.

The fact that the minimum unit and unit range transformations make MRA and SMRA congruent on a context that shares the minimum outcome used to define them (see Figure 2) suggests one escape from these difficulties: Why not use a context-dependent transformation? This approach abandons homoscedasticity with respect to pairs *mc*; hence it is not a strong or strict utility model. This is in fact the approach taken by the contextual utility model in section 5, and it can be given both good psychological motivation and firm grounding in terms of Pratt's (1964) main theorem. First, however, it is worth considering more carefully whether a well-chosen form of agent heteroscedasticity might be adequate. We will see that it is not.

4. Construction and critique of an agent heteroscedasticity approach

Figure 4 shows two concave CRRA utility functions, for two agents *a* and *b*: A square-root (ρ^b = 0.5) CRRA utility function, and a "near-log" (ρ^a = 0.99) CRRA utility function; clearly $a \underset{mra}{\succ} b$. Consider the MPS pair $S_{2,-0} = (0,1,0)$ and $T_{2,-0} = (1/2,0,1/2)$—a sure 100 versus even chances of either 50 or 150. For $a \underset{mra}{\succ} b$ to imply $a \underset{smra}{\succ} b$ under strong utility, the *V*-difference in this pair with ρ^b = 0.99 must exceed that with ρ^a = 0.50: Otherwise the more risk-averse utility function will not imply a higher probability of choosing the safer $S_{2,-0}$. A special transformation that equates first derivatives of utility functions at the intermediate outcome 100 allows Figure 4 to be drawn that way, nesting the *V*-difference with ρ^b = 0.5 inside the *V*-difference with ρ^a = 0.99.[7] The local absolute risk aversion measure $-u''(z)/u'(z)$ may be thought of as measuring concavity relative to slope at *z*: With slopes equated at *z* = 100, Figure 4 does this visually.

Reflection on Figure 4 suggests that for $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$ on some collection of contexts, we might succeed with a transformation of *u*(*z*) that equalizes first derivatives $u'(z)$ of all utility functions at some $\dot{z}$ sufficiently greater than $\bar{z}_1 \equiv \max_c\{z_{1c}\}$, the maximum of the minimum outcomes found in any of several contexts *c* for which we have choice data. Without this, utility functions become arbitrarily flat over one or more contexts as the coefficient of relative risk

[7] In Figure 4, a "level" is also chosen for each utility function such that *u*(100) = 100 for both utility functions. This choice is irrelevant to any strong utility *V*-difference (it differences out) but it allows for easy visual comparison. The matching of first derivatives at z = 100 is the important move. Similar remarks apply to Figure 6.

aversion get large. This in turn implies that all V-differences between lotteries on such contexts approach zero as coefficients of relative risk aversion get large and, hence, that all strong or strict utility choice probabilities approach 0.5 for sufficiently great relative risk aversion on such contexts. The first derivative of the CRRA basic transformation is $u_b'(z) = z^{-\rho}$: Multiplying the basic transformation by $\dot{z}^{\rho}$, or (what is the same thing) $e^{\alpha\rho}$ where $\alpha = \ln(\dot{z})$ for some $\dot{z} > \bar{z}_1$, we give all CRRA utility functions a common unit slope at $\dot{z} = e^{\alpha}$. Therefore, call $u_{sc}(z \mid \rho) =$ $e^{\alpha\rho} z^{1-\rho}/(1-\rho)$ an SMRA-compatible CRRA transformation. Figure 4 is drawn using this transformation, with $\alpha = \ln(100) = 4.61$ giving all CRRA functions a unit slope at $z = 100$.

If we write the CRRA EU V-difference using the SMRA-compatible transformation, the term $e^{\alpha\rho}$ factors out of the V-difference. We then have the strong utility model

$$P_{mc}^{n} = H\left(\lambda e^{\alpha\rho^{n}} \Delta_b V_{mc}(\rho^{n})\right). \tag{4.1}$$

Suppose, then, that we estimate a CRRA EU strong utility model one subject at a time, using the basic transformation: This would be the model $P_{mc}^{n} = H[\lambda^{n} \Delta_b V_{mc}(\rho^{n})]$. Let $\hat{\lambda}^{n}$ and $\hat{\rho}^{n}$ be our estimates for each subject n. If (4.1) is correct, $\ln(\lambda^{n}) \equiv \ln(\lambda) + \alpha\rho^{n}$, and we should therefore expect a linear relationship between $\ln(\hat{\lambda}^{n})$ and $\hat{\rho}^{n}$. If this linear relationship has a slope α significantly greater than $\ln(\dot{z})$, then $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$ in the population from which subjects are drawn on all contexts found in the experiment that generated the data.

Hey's (2001) remarkable data set contains 500 lottery choices per subject, for fifty-three subjects. This allows one to estimate model parameters separately for each subject n with relative precision. The specification of choice probabilities actually used for this individual subject estimation adds a few small features for certain peripheral reasons. It is

$$P_{mc}^n = (1-\delta_{mc})\{(1-\omega^n)H(\lambda^n\Delta_b V_{mc}(\rho^n)) + \omega^n/2\} + \delta_{mc}(1-\omega^n/2)\,. \tag{4.2}$$

As is the case with several existing experimental data sets, Hey's experiment contains a small number of pairs *mc* in which S_{mc} first-order stochastically dominates R_{mc}. Letting Ω_{fosd} be the set of all such FOSD pairs, let $\delta_{mc} = 1\ \forall\ m \in \Omega_{\text{fosd}}$ ($\delta_m = 0$ otherwise). It is well-known that strong and strict utility do a poor job of predicting the rareness of violations of FOSD (Loomes and Sugden 1998). Equation (4.2) takes account of this, yielding $P_{mc}^n = 1-\omega^n/2\ \forall\ mc \in \Omega_{\text{fosd}}$.

The new parameter ω^n is a tremble probability. Some randomness of observed choice has been thought to arise from attention lapses or simple responding mistakes that are independent of pairs *mc*, and adding a tremble probability as above takes account of this. It also provides a convenient way to model the low probability of FOSD violations. Moffatt and Peters (2001) find significant evidence of positive tremble probabilities even in data from experiments where there are no FOSD pairs, such as Hey and Orme (1994). The specification (4.2) follows Moffatt and Peters in assuming that "tremble events" occur with probability ω^n independent of pairs *mc* and, in the event of a tremble, that choices of S_{mc} or T_{mc} from pair *mc* are equiprobable.

Figure 5 plots maximum likelihood estimates $\ln(\hat{\lambda}^n)$ and $\hat{\rho}^n$ from equation (4.2) for each of Hey's (2001) fifty-three subjects, along with a robust regression line between them. It does appear to be a remarkably linear relationship, as suggested by the heteroscedastic form $\ln(\lambda^n) \equiv \ln(\lambda) + \alpha\rho^n$. The robust regression coefficient is $\hat{\alpha} = 4.24$, with a 90% confidence interval [4.11,4.38].[8] There are sixteen MPS pairs in Hey's experimental design: Computational methods easily find the minimum value $\dot{\underline{z}} > \bar{z}_1 \equiv \max_c\{z_{1c}\} = 50$ for those MPS pairs, such that

[8] The robust regression technique used is due to Yohai (1987), and deals with outliers in both the dependent and independent variable. This is appropriate here, since both are estimates.

any $\alpha > \ln(\dot{\underline{z}})$ allows MRA to imply SMRA for those MPS pairs.[9] This minimum value is about $\dot{\underline{z}} = 59.7$, so we need $\alpha > \ln(59.7) = 4.09$ for $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$. As can be seen, the point estimate $\hat{\alpha}$ =4.24 does indeed allow it, and the 90% confidence interval for $\hat{\alpha}$ just allows one to reject (at 5%) the directional null $\alpha \leq 4.09$ in favor of the alternative that $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$ amongst Hey's subjects for all MPS pairs in all of the experiment's contexts. Results are essentially similar if, instead, CRRA RDU is estimated using Prelec's (1998) one-parameter weighting function.

There is, however, an obvious difficulty with this agent heteroscedasticity solution. Suppose we invited Hey's fifty-three subjects back for another experimental session in which they made choices from lottery pairs on a new fifth context $c^* = (100,150,200)$. If we believe the relationship shown in Figure 5 is a fixed one for these subjects, independent of any outcome context they might face, then $a \underset{mra}{\succ} b \not\Rightarrow a \underset{smra}{\succ} b$ on the new context c^*. The estimated upper confidence limit on $\hat{\alpha}$, 4.38, implies a value of $\dot{z} = 79.8 < 100 = z_{1c^*}$ (since $e^{4.38} = 79.8$). Therefore, $\lambda e^{4.38\rho}\Delta_b V_{mc^*}(\rho)$ will converge to zero for sufficiently large ρ and so cannot be monotonically increasing in ρ on the new context c^*.

Figures 6 and 7 illustrate the difficulty. Figure 6 shows six CRRA utility functions (for ρ = 0.25, 0.5, 1, 2, 4 and 8), all drawn using the SMRA-compatible transformation and α = 4.24, the point estimate from the robust regression shown in Figure 5, so that all slopes are unity at $e^{4.24} = 69.41$. While the relative curvature of these utility functions is substantial across the outcome range [0,150] of Hey's experiment, this is not true on the new context (100,150,200). There, increasing ρ eventually results in arbitrarily flat utility functions with arbitrarily small

[9] Starting at $\alpha = \ln(\bar{z}_1) = \ln(50)$, α is incremented in small steps until $e^{\alpha\rho^n}\Delta_b V_{mc}(\rho^n)$ is monotonically increasing in ρ (also checked by computational methods at each value of α) for all MPS pairs *mc* in Hey's design.

differences $u(200)-u(100)$, and this eventually implies arbitrarily small *V*-differences on the new context for sufficiently large ρ. Figure 7 graphs $e^{\alpha\rho}\Delta_b V_{1,c^*}(\rho)$, the CRRA EU *V*-difference in the MPS pair $S_{1,c^*}=(0,7/8,1/8)$ and $T_{1,c^*}=(3/8,1/8,4/8)$, at the lower and upper confidence limits of α from the robust regression, as well as at $\alpha = 4.61 = \ln(100)$ and $\alpha = 4.79 = \ln(120)$. As can be seen, this function is not monotone increasing in ρ until we choose α sufficiently greater than the natural logarithm of the minimum outcome on the new context c^*.

This problem is entirely general. Estimate or choose any finite value of α you wish: A new context c^* with $z_{1c^*} > e^{\alpha}$ always exists such that $a \underset{mra}{\succ} b \nRightarrow a \underset{smra}{\succ} b$ on that new context, given that value of α. Although a suitably chosen α makes the transformation $u_{sc}(z \mid \rho) = e^{\alpha\rho} z^{1-\rho}/(1-\rho)$ "SMRA compatible" on a <u>given</u> collection of contexts, it cannot be SMRA-compatible on <u>all</u> contexts with arbitrarily large values of z_{1c}. Of course, one could make α context-dependent, but this means abandoning strong and strict utility, since there will then be heteroscedasticity with respect to contexts. If this is what we must do in order to make $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$, it is better to approach it in a more basic theoretical way. This is what the contextual utility model does.

5. Contextual utility

Psychological motivation for contextual heteroscedasticity has its origin in signal detection and stimulus discrimination experiments. In this literature, stimulus categorization errors are known to increase with the subjective range taken by the stimulus or signal. For instance, Pollack (1953) and Hartman (1954) presented subjects with tones equally spaced over a range of tones. The <u>range</u> of tones used varies across subjects, but all subjects encounter specific <u>target tones</u>. Confusion of target tones is more common when the overall range of tones encountered is wider.

Such observations gave rise to models of stimulus categorization and discrimination error predicting that classification error variance increases with the subjective range of the stimulus (Parducci 1965, 1974; Holland 1968; Gravetter and Lockhead 1973). In fact, a rough proportionality between subjective stimulus range and the standard deviation of latent error seemed descriptive of much data from categorization experiments, though some of the formal models allowed for some deviation from this (e.g. Holland 1968 and Gravetter and Lockhead 1973). In categorization experiments, where stimuli are presented one at a time and subjects' task is to assign the stimulus to a category, the subjective range of the stimulus was usually taken to be determined (after a period of adaptation) by the whole range of stimuli presented over the course of the experiment—what one might call the "global context" of the stimulus.

The contextual utility model borrows the idea that the standard deviation of evaluative noise is proportional to the subjective range of stimuli from this literature on the perception of stimuli. However, being a model of choice from lottery pairs rather than a model of categorization of singly presented stimuli, it assumes that choice pairs create their own "local context" or idiosyncratic subjective stimulus range, in the form of the range of outcome utilities found in the pair. We may think of agents as perceiving lottery value on context *c* relative to the range of possible lottery values on context *c*. Letting $V(z|\beta^n)$ be agent *n*'s value of a degenerate lottery that pays *z* with certainty, the subjective stimulus range for agent *n* on any context *c* is assumed proportional to $V(z_{3c}|\beta^n)-V(z_{1c}|\beta^n)$. Assume that evaluative noise is proportional to this subjective stimulus range. For non-FOSD pairs on a three-outcome context, and ignoring any tremble for clarity, contextual utility choice probabilities are:

$$P_{mc}^n = H\left(\lambda \frac{V(S_{mc}|\beta^n)-V(T_{mc}|\beta^n)}{V(z_{3c}|\beta^n)-V(z_{1c}|\beta^n)}\right). \tag{5.1}$$

Under RDU, this may be rewritten as

$$P_{mc}^{n} = H\left(\lambda \frac{(ws_{m1} - wt_{m1})u^{n}(z_{1c}) + (ws_{m2} - wt_{m2})u^{n}(z_{2c}) + (ws_{m3} - wt_{m3})u^{n}(z_{3c})}{u^{n}(z_{3c}) - u^{n}(z_{1c})} \right). \tag{5.2}$$

Since $(ws_{m1} - wt_{m1}) \equiv -(ws_{m2} - wt_{m2}) - (ws_{m3} - wt_{m3})$, (5.2) may be rewritten as

$$P_{mc}^{n} = H\big(\lambda[(ws_{m2} - wt_{m2})\upsilon_{c}^{n}(z_{2c}) + (ws_{m3} - wt_{m3})]\big), \tag{5.3}$$

where $\upsilon_{c}^{n}(z) \equiv [u^{n}(z) - u^{n}(z_{1c})]/[u^{n}(z_{3c}) - u^{n}(z_{1c})]$ can be recognized as a context-specific unit range transformation of agent *n*'s utility function. We may therefore view contextual utility as employing a "contextual unit range transformation" of the utility function.

Suppose that pair *mc* in model (5.3) is an MPS pair. Pratt's (1964, p. 128) Theorem 1 quickly shows that contextual utility guarantees that $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$ (holding weighting functions and λ constant across agents) on all three-outcome contexts, as hinted by Figure 2. Here I quote the relevant parts of Pratt's theorem, with notation modified to fit mine (all italics in original):

> Theorem 1: *Let* $r^{n}(z)$ [equal to $-u''(z)/u'(z)$ for utility function u^{n}] *be the local risk aversion…corresponding to the utility function* u^{n}*, n = a,b. Then the following conditions are equivalent, in either the strong form (indicated in brackets), or the weak form (with the bracketed material omitted).*
>
> (a) $r^{a}(z) \geq r^{b}(z)$ *for all z* [*and* > *for at least one z in every interval*]…
>
> (e) $\dfrac{u^{a}(z_{3c}) - u^{a}(z_{2c})}{u^{a}(z_{2c}) - u^{a}(z_{1c})} \leq [<] \dfrac{u^{b}(z_{3c}) - u^{b}(z_{2c})}{u^{b}(z_{2c}) - u^{b}(z_{1c})}$ *for all* z_{1c}, z_{2c}, z_{3c} *with* $z_{1c} < z_{2c} < z_{3c}$.

Adding unity to both sides of (e) in the form $[u^{n}(z_{2c}) - u^{n}(z_{1c})]/[u^{n}(z_{2c}) - u^{n}(z_{1c})]$, and then taking reciprocals of both sides, we get the following equivalence from the theorem:

$$r^{a}(z) > r^{b}(z) \forall z \in [z_{1c}, z_{3c}] \Leftrightarrow \frac{u^{a}(z_{2c}) - u^{a}(z_{1c})}{u^{a}(z_{3c}) - u^{a}(z_{1c})} > \frac{u^{b}(z_{2c}) - u^{b}(z_{1c})}{u^{b}(z_{3c}) - u^{b}(z_{1c})} \forall z_{2c} \in (z_{1c}, z_{3c}). \tag{5.4}$$

Since $\upsilon_c^n(z_{2c}) = [u^n(z_{2c}) - u^n(z_{1c})]/[u^n(z_{3c}) - u^n(z_{1c})]$, equation (5.4) shows that

$$r^a(z) > r^b(z) \forall z \in [z_{1c}, z_{3c}] \Leftrightarrow \upsilon_c^a(z_{2c}) > \upsilon_c^b(z_{2c}) \forall z_{2c} \in (z_{1c}, z_{3c}) . \quad (5.5)$$

Equation (5.3) shows that the RDU V-difference between S_{mc} and T_{mc} under the contextual unit range transformation is $(ws_{m2} - wt_{m2})\upsilon_c^n(z_{2c}) + (ws_{m3} - wt_{m3})$, which is obviously increasing in $\upsilon_c^n(z_{2c})$ for all MPS pairs since $(ws_{m2} - wt_{m2}) > 0$ in these pairs, as shown in (1.4). Therefore, by equation (5.5), it is increasing in local risk aversion $r^n(z)$ as well. As a result, the contextual utility model easily implies what strong and strict utility cannot: That $a \underset{mra}{\succ} b \Rightarrow a \underset{smra}{\succ} b$ on all three-outcome contexts when λ and weighting functions are held constant across agents.

Some agents may be less precise than others: Precision λ may vary across agents. The following two propositions refine the result to allow this (proofs are obvious and so omitted):

Proposition 1: *Consider two EU agents such that* $\lambda^a \geq \lambda^b$ *and* $r^a(z) > r^b(z) \geq 0$ *for all z. Then in an EU contextual utility model,* $a \underset{smra}{\succ} b$*. Put differently,* $a \underset{mra}{\succ} b$ *and* $\lambda^a \geq \lambda^b \Rightarrow$ $a \underset{smra}{\succ} b$ *under an EU contextual utility model.*

The restriction to EU matters since RDU agents may prefer the riskier lottery in some MPS pairs: When they do, they have a negative V-difference in the pair. Larger value of λ will magnify that and possibly offset greater risk aversion in Pratt's sense. For RDU, then, we confine attention to MPS pairs where the less risk-averse agent b (in Pratt's sense) prefers the safe lottery:

Proposition 2: *Consider two RDU agents such that* $\lambda^a \geq \lambda^b$, $r^a(z) > r^b(z) \geq 0$ *for all z, and* $w^a(q) \equiv w^b(q)$*. Then in an RDU contextual utility model,* $P_{mc}^a > P_{mc}^b$ *for all MPS pairs in which* $P_{mc}^b \geq 0.5$.

Note that contextual utility does not imply that safe choices become certain as risk aversion (in Pratt's sense) becomes infinite. This is easily seen by noting that $\upsilon_c^n(z_{2c}) \to 1$ as $r^n(z) \to \infty$, in which case equation (5.3) with agent-dependent precision λ^n becomes

$$P_{mc}^n = H\big(\lambda^n[w^n(s_{m2}+s_{m3}) - w^n(t_{m2}-t_{m3})]\big). \tag{5.6}$$

Since $w^n(s_{m2}+s_{m3}) - w^n(t_{m2}+t_{m3})$ is obviously finite, contextual utility does not imply that $P_{mc}^n \to 1$ as $r^n(z) \to \infty$ for all MPS pairs; it just implies that P_{mc}^n is increasing in $r^n(z)$ for given λ^n and $w^n(q)$. Of course, there will always exist a finite λ^n such that P_{mc}^n approaches certainty to any specified degree as $r^n(z)$ becomes large.

When combined with either a constant absolute risk aversion (CARA) utility function $u(z) = -e^{-rz}$ or a CRRA utility function, contextual utility implies some theoretically attractive invariance properties of choice probabilities. Call $c+k = (z_{1c}+k, z_{2c}+k, z_{3c}+k)$ and $kc = (kz_{1c}, kz_{2c}, kz_{3c})$ additive and proportional shifts of context $c = (z_{1c}, z_{2c}, z_{3c})$, respectively. Since $u(z+k) = -e^{-r(z+k)} = -e^{-rk}e^{-rz}$ for CARA, we have

$$\upsilon_{c+k}^n(z+k) \equiv \frac{(-e^{-rk}e^{-rz} + e^{-rk}e^{-rz_{1c}})}{(-e^{-rk}e^{-rz_{3c}} + e^{-rk}e^{-rz_{1c}})} \equiv \frac{(-e^{-rz} + e^{-rz_{1c}})}{(-e^{-rz_{3c}} + e^{-rz_{1c}})} \equiv \upsilon_c^n(z). \tag{5.7}$$

Equation (5.3) then implies that $P_{m,c+k}^n \equiv P_{mc}^n$. Similarly, since $u(kz) = (zk)^{1-\rho} = k^{1-\rho}z^{1-\rho}$ for CRRA, we have

$$\upsilon_{kc}^n(kz) \equiv \frac{(k^{1-\rho}z^{1-\rho} - k^{1-\rho}z_{1c}^{1-\rho})}{(k^{1-\rho}z_{3c}^{1-\rho} - k^{1-\rho}z_{1c}^{1-\rho})} \equiv \frac{(z^{1-\rho} - z_{1c}^{1-\rho})}{(z_{3c}^{1-\rho} - z_{1c}^{1-\rho})} \equiv \upsilon_c^n(z). \tag{5.8}$$

Equation (5.3) then implies that $P_{m,kc}^n \equiv P_{mc}^n$. That is, contextual utility implies that choice probabilities in pairs are invariant to an additive (proportional) context shift given a CARA (CRRA) utility function. This property echoes what is well-known about structural CARA and

CRRA preferences, namely that their preference directions are invariant to additive and proportional context shifts, respectively.[10]

For sets of pairs on a single context, contextual utility will share all properties of strong utility models, such as SST (Luce and Suppes 1965) and simple scalability (Tversky and Russo 1969), since contextual utility is observationally identical to a strong utility model with agent heteroscedasticity on a single context. However, because contextual utility is heteroscedastic across contexts, it will violate SST and simple scalability for sets of pairs on several distinct contexts: In general it only obeys moderate stochastic transitivity or MST (see Appendix A). This is a descriptive bonus: Contextual utility will explain well-known violations of simple scalability such as the Myers effect (Myers and Sadler 1960) in much the same way other heteroscedastic models such as decision field theory do (Busemeyer and Townsend 1993).[11]

Others have posited pair heteroscedasticity in discrete choice under risk. Both Hey (1995) and Buschena and Zilberman (2000) investigated several heteroscedastic forms for econometric reasons and/or theoretical reasons based on similarity relations.[12] Though it is not Blavatskyy's (2007) central innovation, part of his heteroscedastic form is precisely that posited by contextual utility and Blavatskyy's reason for doing this is unadorned (but good) intuition. Busemeyer and Townsend's (1993) decision field theory also produces a complex form of heteroscedasticity that varies with outcome utilities and probabilities. Its logic is stochastic sampling of outcome

[10] The logarithmic *V*-difference strict utility form (2.5) and random preference models share these two shift invariance properties with contextual utility, but strong utility does not. Contextual utility and strong utility EU models share a different property which might be called the "false common ratio effect" or FCRE, while random preference EU models do not have this property. See Loomes (2005, pp. 303-305) for a lucid discussion of the manner in which strong utility EU produces the FCRE (the reasoning is identical for contextual utility) and why random preference EU does not. See Wilcox (2007a) for an extensive comparison of properties of combinations of EU and RDU structures with various stochastic models.

[11] Busemeyer and Townsend (1993) explain simple scalability, how the "Myers effect" violates it, and the logic behind decision field theory's explanation of the Myers effect (contextual utility gives a very similar explanation).

[12] A class of heteroscedastic models described by Carroll and De Soete (1991) handle certain kinds of similarity: See Wilcox (2007a) for an application of their "wandering vector model" to choice under risk.

utilities, partially guided by outcome probabilities: Put differently, the heteroscedasticity arises from computational reasons. Contextual utility's relatively unique and distinguishing feature is that it arrives at contextual heteroscedasticity by interrogating the logic of the relationship between MRA and SMRA in latent variable models, rather than a computational logic (though contextual utility has some empirical psychological grounding as well). It is surprising and interesting that similar (though by no means identical) forms of pair heteroscedasticity can be the conclusion of such different theoretical approaches to the issue of stochastic choice under risk.

6. An empirical comparison of contextual utility and some competitors

The question naturally arises: Which stochastic model, when combined with EU or RDU, actually explains and predicts binary choice under risk best? To answer this question, we need a suitable data set. Strong utility and contextual utility are simple reparameterizations of one another for any one context. Therefore, data from any experiment where no subject makes choices from pairs on several distinct contexts (e.g. Loomes and Sugden 1998) are not suitable. The experiment of Hey and Orme (1994), hereafter HO, is suitable since all subjects make choices from pairs on four distinct contexts.

The HO experiment has another desirable design feature that is unique among experiments with multiple contexts: The same pairs of probability distributions are used to construct the lottery pairs on all four of its contexts. Therefore, when models fail to explain choices across multiple contexts in the HO data, we cannot attribute this to other differences in lottery pairs across contexts: The failure must be due to the model's generalizability across contexts.

The HO experiment builds lotteries from four equally spaced non-negative outcomes including zero, in increments of 10 UK pounds. Letting "1" represent 10 UK pounds, the experiment uses the outcome vector (0,1,2,3). Exactly one-fourth of observed lottery choices in the HO data are choices from pairs on each of the four possible three-outcome contexts from this vector. In keeping with past notation, these four contexts are denoted by $c \in \{-0,-1,-2,-3\}$, where $-0 = (1,2,3)$, $-1 = (0,2,3)$, $-2 = (0,1,3)$ and $-3 = (0,1,2)$.

HO estimate a variety of structures combined with strong utility, and do this individually—that is, they estimate each structure separately for each subject. Additionally, for all structures that specify a utility function on outcomes $u(z)$, HO take a nonparametric approach to the utility function. Letting $u_z^n \equiv u^n(z)$, HO set $u_0^n = 0$ and $\lambda = 1$, and estimate u_1^n, u_2^n and u_3^n directly, allowing the utility function $u(z)$ to take on arbitrary shapes across the outcome vector (0,1,2,3). (The form of strong utility models and the affine transformation properties of $u(z)$ imply that just three of the five parameters λ, u_0^n, u_1^n, u_2^n and u_3^n are identified.) HO found that estimated utility functions overwhelmingly fall into two classes: Concave utility functions, and inflected utility functions that are concave over the context (0,1,2) but convex over the context (1,2,3). The latter class accounts for thirty to forty percent of subjects (depending on the structure estimated). Because of this, I follow HO in avoiding a simple parametric functional form such as CARA or CRRA that forces concavity or convexity across the entire outcome vector (0,1,2,3), instead adopting their nonparametric treatment of utility functions in strong, strict and contextual utility models. However, I will instead set $u_0^n = 0$ and $u_1^n = 1$, and estimate λ, u_2^n and u_3^n.

To account for heterogeneity of subject behavior, I part company with HO's individual estimation and use a random parameters approach for estimation, as done by Loomes, Moffatt and Sugden (2002) and Moffatt (2005). Individual estimation does avoid distributional assumptions made by random parameters methods, and seems well-behaved at HO's sample sizes when we confine attention to "in-sample" comparisons of model fit, as HO did. However, Monte Carlo analysis shows that with finite samples of the HO size, comparisons of the out-of-sample predictive performance of stochastic models can be extremely misleading with individual estimation and prediction (Wilcox 2007b). I want to address both the in-sample fit and out-of-sample predictive performance of models, so this issue matters here. Proper accounting for heterogeneity is crucial when comparing structural models of discrete choice under risk (Wilcox 2007a). With both of these issues in mind, I choose a random parameters estimation approach.[13]

The HO experiment allowed subjects to express indifference between lotteries. HO model this with an added "threshold of discrimination" parameter within a strong utility model. An alternative parameter-free approach, and the one I take here, treats indifference in a manner suggested by decision theory, where the indifference relation $S_{mc} \sim^n T_{mc}$ is defined as the intersection of two weak preference relations, i.e. "$S_{mc} \succeq^n T_{mc} \cap T_{mc} \succeq^n S_{mc}$." This suggests treating indifference responses as two responses in the likelihood function—one of S_{mc} being chosen from mc, and another of T_{mc} being chosen from mc—but dividing that total log likelihood by two since it is really based on just one independent observation. Formally, the definite choice of S_{mc} adds $\ln(P_{mc})$ to the total log likelihood; the definite choice of T_{mc} adds $\ln(1-P_{mc})$ to that total; and indifference adds $[\ln(P_{mc})+\ln(1-P_{mc})]/2$ to that total. See also Papke and Wooldridge (1996) and Andersen et al. (2007) for related justifications of this approach.[14]

The random preference model also appears in both contemporary experimental and theoretical work on stochastic discrete choice under risk (Loomes and Sugden 1995, 1998; Carbone 1997; Loomes, Moffatt and Sugden 2002; Gul and Pesendorfer 2006). Econometrically,

[13] This does cost some extra parameters—eight in all for EU specifications, and 11 in all for RDU specifications. This allows for salient covariation of risk and precision parameters in the sampled population. My own opinion is that guarding against aggregation biases is well worth this cost, especially in a large data set like HO with 12000 observations. See Appendix C and Wilcox (2007a) for details of the specifications.

[14] Note that indifference responses are rare overall in the HO data (about 2.7% of all responses) and that these are concentrated amongst a relatively small number of subjects.

the random preference model views stochastic choice as arising from randomness of structural parameters. We think of each agent *n* as having an urn filled with structural parameter vectors β^n. Following Carbone (1997) for instance, we may set $u_0^n = 0$ and $u_1^n = 1$ and think of EU with random preferences as a situation where $\beta^n = (u_2^n, u_3^n)$, with $u_2^n > 1$ and $u_3^n > u_2^n$. Each agent *n* has an urn filled with such utility vectors (u_2^n, u_3^n). At each trial of any pair *mc*, an agent draws one of these vectors from her urn (with replacement) and uses it to calculate both $V(S_{mc} \mid \beta^n)$ and $V(T_{mc} \mid \beta^n)$ without error, choosing S_{mc} iff $V(S_{mc} \mid \beta^n) \geq V(T_{mc} \mid \beta^n)$. Let $F_\beta(x \mid \alpha^n)$ be the joint c.d.f. of β in agent *n*'s "random preference urn," conditioned on some vector α^n of parameters determining the shape of the distribution F_β. Then under random preferences,

$$P_{mc}^n = \Pr\big(V(S_{mc} \mid \beta^n) - V(T_{mc} \mid \beta^n) \geq 0 \mid F_\beta(x \mid \alpha^n)\big). \tag{6.1}$$

Loomes, Moffatt and Sugden (2002) show how to specify a random preferences RDU model for pairs on a single context. However, specification of a random preference RDU model across multiple contexts is more difficult. Building both on Loomes, Moffatt and Sugden and certain insights of Carbone (1997), Appendix B shows how this may be done for the three contexts −0, −2 and −3, but also shows why further extending random preferences to cover the other context −1 is not a transparent exercise for the RDU structure. For this reason, I confine all of my estimations to choices in the HO data on the contexts −0, −2 and −3, where a random preference RDU model can clearly be compared to strong, strict and contextual utility RDU models.

Appendix C illustrates the random parameters specification and estimation in detail for the EU structure with strong utility; see Wilcox (2007a) for detailed exposition of all specifications and their estimation. I perform two different kinds of comparisons of model fit. The first kind (very common in this literature) are "in-sample fit comparisons." Models are estimated on all three of the HO contexts (–0, −2 and −3) and the resulting log likelihoods for each model across all three contexts are compared.

The second kind of comparison, which is rare in this literature, compares the "out-of-sample" fit of models—that is, their predictive performance.[15] For these comparisons, models are estimated on just the two HO contexts −2 and −3, that is contexts (0,1,3) and (0,1,2), and these estimates are used to predict choice probabilities and calculate log likelihoods of observed choices on HO context –0, that is context (1,2,3). This is something more than a simple out-of-sample prediction, which could simply be a prediction to new choice trials of the same pairs (and hence contexts) used for estimation: It is additionally an "out-of-context" prediction.[16]

This particular kind of out-of-sample fit comparison may be quite difficult in the HO data. Relatively safe choices are the norm in contexts −2 and −3 of the HO data: The mean proportion of safe choices made by HO subjects in these contexts is 0.764, and at the individual level this proportion exceeds ½ for seventy of the eighty subjects. But relatively risky choices are the norm in context –0 of the HO data: The mean proportion of safe choices there is just 0.379, and falls short of ½ for fifty-eight of the eighty subjects. Recall that we cannot attribute this to differences in the probability vectors that make up the lottery pairs in the different HO contexts, since the HO experiment holds this constant across contexts. This out-of-sample prediction task is going to be difficult: From largely safe choices in the "estimation contexts" −2 and −3, the models need to predict largely risky choices in the "prediction context" –0.

---

[15] The HO design presents the same 100 pairs (25 on each of its four contexts) to a fixed group of subjects on two separate days. This is a panel of subjects with variation in pairs, contexts and days. Any such panel could be divided up into a "sample" for estimation, and remaining "out-of-sample" observations for predictive evaluation, in different ways. For instance one might divide it up by its time dimension, estimating with one day's choices, and then trying to predict choices made on the other day. Since my interest lies with the different way in which the stochastic models treat choices across contexts, I choose to divide the panel up across its context, rather than time, dimension.
[16] This is exactly Busemeyer and Wang's (2000) distinction between "cross-validation" and "generalization."

Table 1 displays both the in-sample and out-of-sample log likelihoods for the eight models. The top four rows are EU models, and the bottom four rows are RDU models; for each structure, the four rows show results for strong utility, strict utility, contextual utility and random preferences. The first column shows total in-sample log likelihoods, and the second column shows total out-of-sample log likelihoods. Contextual utility always produces the highest log likelihood, whether it is combined with EU or RDU, and whether we look at in-sample or out-of-sample log likelihoods (though the log likelihood advantage of contextual utility is most pronounced in the out-of-sample comparisons). Buschena and Zilberman (2000) and Loomes, Moffatt and Sugden (2002) point out that the best-fitting stochastic model may depend on the structure estimated and offer empirical illustrations of this sensible econometric point. Yet in Table 1 contextual utility is the best stochastic model whether viewed from the perspective of EU or RDU, or from the perspective of in-sample or out-of-sample fit.

Decision theory has been relatively dominated by structural theory innovation over the last quarter century. Table 1 has something to say about this. Examine the in-sample fit column first. Holding stochastic models constant, the maximum log likelihood improvement of switching from EU to RDU is 142.02 (with strict utility), and the improvement is 106.64 for the best-fitting stochastic model (contextual utility). Holding structures constant instead, the maximum improvement in log likelihood associated with changing the stochastic model is 151.48 (with the EU structure, switching from strict to contextual utility), but this is atypical: Omitting strict utility models, which have an unusually poor in-sample fit, the maximum improvement is 51.28 (with the EU structure, switching from strong to contextual utility) and otherwise no more than half that. Therefore, except for the especially poor strict utility fits, in-sample comparisons make choice of a stochastic model appear to be a sideshow relative to choice of a structure.

This appearance is reversed when we look at out-of-sample predictive power. Looking now at the out-of-sample fit column, notice first that under strict utility, RDU actually fits worse than EU does. We should be getting the impression by now, however, that strict utility is an unusually poor performer, so let us henceforth omit it from consideration. Among the remaining three stochastic models, the maximum out-of-sample fit improvement associated with switching from EU to RDU is 21.19 (for contextual utility). Holding structures constant instead, the maximum out-of-sample fit difference between the stochastic models (again omitting strict utility) is 113.39 (for RDU, switching from strong to contextual utility). In out-of-sample prediction, then, structures play Rosencrantz and Guildenstern to the stochastic model Hamlet. Decision research seems heavily preoccupied with explanation. But these results suggest that those who have more interest in prediction may want to think more about stochastic models, as repeated urged by Hey (Hey and Orme 1994; Hey 2001; Hey 2005) and suggested by Ballinger and Wilcox (1997).

Table 2 reports a formal comparison of stochastic models, conditional on each structure. Let $\tilde{D}^n$ be the difference between the estimated log likelihoods (in-sample or out-of-sample) from a pair of models, for subject *n*. Vuong (1989) shows that asymptotically, a *z*-score based on the $\tilde{D}^n$ follows a normal distribution under the null that two non-nested models are equally close to the truth (neither model needs to be the true model). The statistic is $z = \sum_{n=1}^{N} \tilde{D}^n / (\tilde{s}_D \sqrt{N})$, where $\tilde{s}_D$ is the sample standard deviation of the $\tilde{D}^n$ across subjects *n* (calculated without adjustment for a degree of freedom) and *N* is the number of subjects. Table 2 reports these *z*-statistics and *p*-values against the null of equally good fit, with a one-tailed alternative that the directionally better fit is significantly better. While contextual utility is always directionally better than its competitors, no convincingly significant ordering of the stochastic models emerges from the in-sample comparisons shown in the left half of Table 2, though strict utility is clearly significantly worse than the other three stochastic models. Contextual utility shines, though, in the out-of-sample fit comparisons in the right half of Table 2, regardless of whether the structure is EU or RDU, where it beats the other three stochastic models with strong significance.

7. Conclusions

While most of the scholarly conversation about decision under risk concerns its structure, there is resurgent interest in the stochastic part of decision under risk. This has been driven both by theoretical questions and empirical findings. Theoretically, some or all of what passes for "an anomaly" relative to some structure (usually EU) can sometimes be attributed to stochastic models rather than the structure in question (Wilcox 2007a). This is an old point, stretching back at least to Becker, DeGroot and Marschak's (1963a,1963b) observation that violations of betweenness are precluded by some stochastic versions of EU (random preferences) but predicted by other stochastic versions of EU (strong utility). But this general concern has been resurrected by many writers; Loomes (2005), Gul and Pesendorfer (2006) and Blavatskyy (2007) are just three relatively recent (but very different) examples. Empirically, we know from a (now quite large) body of experimental evidence on retest reliabilities that the stochastic part of decision under risk is surprisingly large. Additionally, a "new structural econometrics" has emerged over the last decade in which structures like EU, RDU and CPT are combined with some stochastic model for the purpose of estimating structural risk parameters.

I have interrogated the empirical meaning of structural parameters in econometric models of risky choice. I find that in strong and strict utility models, structural parameters meant to represent the degree of risk aversion in Pratt's (1964) sense cannot order agents according to a theoretically attractive definition of the relation "stochastically more risk averse" across all choice contexts. I conclude that strong and strict utility are deeply troubled for the econometrics of discrete choice under risk. Contextual utility eliminates this trouble without introducing extra parameters. In the Hey and Orme (1994) data set, its in-sample explanatory performance exceeds that of strong and strict utility models and the random preference model; and its out-of-sample predictive performance is significantly the best of this particular collection of stochastic models.

I regard stochastic choice as the oldest and most robust fact of choice under risk, and believe that serious interpretive errors can occur when the implications of stochastic choice models are ignored. I have shown that when choice is stochastic, a globally coherent notion of greater risk aversion necessarily implies the existence of certain context effects. Decision research views some contextual labilities of choice as failures of rationality. Surely some are, but not all are: Some may be the prosaic consequence of a sensible stochastic model that makes global sense of SMRA. For instance, the Myers effect (Myers and Sadler 1960) appears to be a reversal of preferences caused by a change in a standard of comparison. Busemeyer and Townsend (1993) explain why the Myers effect may be no more (or less) than contextual heteroscedasticity.

Contextual utility also has implications for "strength of incentives" in experiments and more generally any mechanism. If contextual utility is correct, the marginal utility difference between taking risky actions X and Y is only part of what governs the strength of perceived incentives: The other important part is the subjective range of available utilities. This implies that scaling up outcomes may be a relatively ineffective means of strengthening incentives. If we can instead redesign an experiment or mechanism to shrink the subjective range of available utilities, while holding marginal utility improvements constant in the neighborhood of a maximum, this may be a more effective way of strengthening subjective optimization incentives.

Appendix A: Stochastic Transitivity Properties of Contextual Utility

Definitions: Let {C,D,E} be any triple of lotteries generating the pairs {C,D}, {D,E} and {C,E}: In a basic triple, all three pairs are basic pairs. A heteroscedastic V-difference latent variable model is $P_{ST} = F[\lambda(V_S - V_T)/\sigma_{ST}]$, where P_{ST} is the probability that S is chosen from pair {S,T}, $V_S \equiv V(S \mid \beta)$ is the structural value V of any lottery S, and σ_{ST} is a noise component specific to pair {S,T}. (Suppress structural parameters β but assume they are fixed, so the discussion is about an individual agent.) Let $\underline{V}_S$ and $\overline{V}_S$ denote the value of degenerate lotteries that pay the minimum and maximum outcomes in lottery S with certainty, respectively. In basic triples, the intervals $[\underline{V}_C, \overline{V}_C]$, $[\underline{V}_D, \overline{V}_D]$ and $[\underline{V}_E, \overline{V}_E]$ must overlap (if not, the outcome ranges of two lotteries in the triple would be disjoint and they would form an FOSD pair). From (5.1), σ_{ST} is the utility range in pair {S,T}, that is $\max(\overline{V}_S, \overline{V}_T) - \min(\underline{V}_S, \underline{V}_T)$, in the contextual utility model.

I make use of Halff's Theorem (Halff 1976): Any heteroscedastic V-difference latent variable model in which pair-specific noise components σ_{ST} obey the triangle inequality across triples of pairs will satisfy moderate stochastic transitivity (MST).

Proposition: Contextual utility obeys MST, but not strong stochastic transitivity (SST), in all basic triples. (Remark: This only rules out triples with glaringly transparent FOSD pairs where all outcomes in one lottery exceed all outcomes in another lottery. See (4.2) for a treatment of FOSD pairs using trembles.

Proof: The utility range in a pair cannot be less than the utility range in either of its component lotteries, so $\sigma_{CD} \geq \overline{V}_C - \underline{V}_C$ and $\sigma_{DE} \geq \overline{V}_E - \underline{V}_E$: Sum to get $\sigma_{CD} + \sigma_{DE} \geq \overline{V}_C - \underline{V}_C + \overline{V}_E - \underline{V}_E$. Since $\{C,D,E\}$ is a basic triple, $[\underline{V}_C, \overline{V}_C]$ and $[\underline{V}_E, \overline{V}_E]$ overlap. Therefore, the utility range in pair $\{C,E\}$ cannot exceed the sum of the utility ranges of its component lotteries C and E: That is, $\sigma_{CE} \leq \overline{V}_C - \underline{V}_C + \overline{V}_E - \underline{V}_E$. Combining the last two inequalities, we have $\sigma_{CD} + \sigma_{DE} \geq \sigma_{CE}$, which is the triangle inequality. By Halff's Theorem, contextual utility obeys MST for all basic triples.

An example suffices to show that contextual utility can violate SST in basic triples. Consider an expected value maximizer. Assume that C, D and E have outcome ranges [0,200], [100,300] and [100,400], respectively, and expected values 162, 160 and 150, respectively. The latent variable in contextual utility is the ratio of a pair's V-difference to the pair's range of possible utilities. In this example, these ratios are 2/300 in pair $\{C,D\}$, 10/300 in pair $\{D,E\}$, and 12/400 = 9/300 in pair $\{C,E\}$. All are positive, implying that all choice probabilities (of the first lottery in each pair) exceed 0.5. But the probability that C is chosen over E will be less than the probability that D is chosen over E, since the latent variable in the former pair (9/300) is less than the latent variable in the latter pair (10/300). This violates SST.

Appendix B: Random Preference RDU Across Multiple Contexts

Using insights of Carbone (1997), I generalize Loomes, Moffatt and Sugden's (2002) technique for single contexts. Like Loomes, Moffatt and Sugden, assume that weighting function parameters are nonstochastic, that is, that the only structural parameters that vary in a subject's "random preference urn" are her outcome utilities. Combining (1.3) and (6.1), we have

$$P_{mc}^{n} = \Pr\left(\sum_{i=1}^{3} W_{mi}(\gamma^{n}) u^{n}(z_{ic}) \geq 0 \mid F_{u}(x \mid \alpha^{n})\right), \text{ where}$$

$$W_{mi}(\gamma^{n}) = w\left(\sum_{j \geq i} s_{mi} \mid \gamma^{n}\right) - w\left(\sum_{j > i} s_{mi} \mid \gamma^{n}\right) - w\left(\sum_{j \geq i} t_{mi} \mid \gamma^{n}\right) + w\left(\sum_{j > i} t_{mi} \mid \gamma^{n}\right). \quad \text{(B.1)}$$

Since $W_{m1}(\gamma^{n}) \equiv -W_{m2}(\gamma^{n}) - W_{m3}(\gamma^{n})$, and assuming strict monotonicity of utility in outcomes so that we may divide the inequality within (B.1) through by $u^{n}(z_{2c}) - u^{n}(z_{1c})$, we have

$$P_{mc}^{n} = \Pr\left(W_{m2}(\gamma^{n}) + W_{m3}(\gamma^{n})[v_{c}^{n} + 1] \geq 0 \mid F_{u}(x \mid \alpha^{n})\right), \text{ where}$$

$$v_{c}^{n} \equiv [u^{n}(z_{3c}) - u^{n}(z_{2c})]/[u^{n}(z_{2c}) - u^{n}(z_{1c})] \ ,$$

$$W_{m2}(\gamma^{n}) = w(s_{m2} + s_{m3} \mid \gamma^{n}) - w(s_{m3} \mid \gamma^{n}) - [w(t_{m2} + t_{m3} \mid \gamma^{n}) - w(t_{m3} \mid \gamma^{n})], \text{ and}$$

$$W_{m3}(\gamma^{n}) = w(s_{m3} \mid \gamma^{n}) - w(t_{m3} \mid \gamma^{n}). \quad \text{(B.2)}$$

Notice that we can view random preferences as based on a context-dependent ratio of differences transformation of utility, namely v_{c}^{n}, as is the case with contextual utility (though the ratio of differences is not the same in the two models). Random preference models will, therefore, also display context dependence. Unlike contextual utility, however, $v_{c}^{n} \in \mathbb{R}^{+}$ is a random variable, containing all choice-relevant stochastic information about the agent's random utility function $u^{n}(z)$ for choices on context *c*. Let $G_{vc}(x \mid \alpha_{c}^{n})$ be the context-specific c.d.f. of v_{c}^{n}, generated by the joint distribution $F_{u}(x \mid \alpha^{n})$ of agent *n*'s random utility vector. Since $t_{m3} - s_{m3} > 0$ in basic pairs, we have $W_{3c}(\gamma^n) = w(t_{m3} | \gamma^n) - w(s_{m3} | \gamma^n) > 0$ for basic pairs.[17] Rewriting (B.2) to make the change of random variables suggested above explicit, we then have

$$P_{mc}^n = \Pr\left(v_c^n \le -[W_{m2}(\gamma^n) + W_{m3}(\gamma^n)] / W_{m3}(\gamma^n) \,|\, G_{vc}(x \,|\, \alpha_c^n)\right), \text{ or}$$

$$P_{mc}^n = G_{vc}\left(\frac{w(s_{m2} + s_{m3} | \gamma^n) - w(t_{m2} + t_{m3} | \gamma^n)}{w(t_{m3} | \gamma^n) - w(s_{m3} | \gamma^n)} \,|\, \alpha_c^n\right). \qquad \text{(B.3)}$$

With equation (B.3) we have arrived where Loomes, Moffatt and Sugden (2002) left things. Choosing some c.d.f. on $\mathbb{R}^+$ for G_{vc}, such as the Lognormal or Gamma distribution, construction of a likelihood function from (B.3) and choice data is straightforward <u>for one context</u> and this is the kind of experimental data Loomes, Moffatt and Sugden had. But when contexts vary in a data set, the method quickly becomes intractable except for special cases. I now work out one such special case. By choosing the utilities of the two lowest outcomes equal to zero and one, respectively, the random utility vector for the four outcomes (0,1,2,3) may be summarized by a random utility vector for just the two highest outcomes, that is the random vector (u_2^n, u_3^n), where $u_2^n > 1$ and $u_3^n > u_2^n$. Let $g_1^n \equiv u_2^n - 1 \in \mathbb{R}^+$ and $g_2^n \equiv u_3^n - u_2^n \in \mathbb{R}^+$ be two underlying random variables generating these random utilities as $u_2^n = 1 + g_1^n$ and $u_3^n = 1 + g_1^n + g_2^n$. A little algebra shows that $v_{-3}^n = g_1^n$, $v_{-2}^n = g_1^n + g_2^n$, $v_{-1}^n = g_2^n / (g_1^n + 1)$, and $v_{-0}^n = g_2^n / g_1^n$.

With the four v_c^n (for the four contexts) expressed this way, we want a joint distribution of g_1^n and g_2^n so that as many of these as possible have tractable parametric distributions. The <u>best</u> choice I am aware of still only works for three of these four expressions. That choice is two independent gamma variates, each with the gamma distribution's c.d.f. $G(x|\phi,\kappa)$, with <u>identical</u> "scale parameter" κ^n but possibly different "shape" parameters ϕ_1^n and ϕ_2^n. Under this choice,

[17] FOSD pair choice probabilities are modeled as tremble events, as in (4.1) and Loomes, Moffatt and Sugden.

$$v_c^n \text{ is distributed...} \quad \begin{cases} \text{Gamma with c.d.f. } G(x \mid \phi_1, \kappa) \text{ for pairs on context } c = -3, \\ \text{Gamma with c.d.f. } G(x \mid \phi_1 + \phi_2, \kappa) \text{ for pairs on context } c = -2, \\ \text{Beta-prime with c.d.f. } B'(x \mid \phi_2, \phi_1) \text{ for pairs on context } c = -0. \end{cases} \quad \text{(B.4)}$$

The "beta-prime" distribution on $\mathbb{R}^+$ is also called a "beta distribution of the second kind" (Aitchison 1963).[18] These assumptions imply a joint distribution of $u_2^n - 1$ and $u_3^n - 1$ known as "McKay's bivariate gamma distribution" and a correlation coefficient $\sqrt{\phi_1^n / (\phi_1^n + \phi_2^n)}$ between u_2^n and u_3^n in subject n's "random preference urn" (Hutchinson and Lai 1990).

An acquaintance with the literature on estimation of random utility models may make these assumptions seem very special and unnecessary. However, theories of choice under risk are special relative to the kinds of preferences that typically get treated in that literature. Consider the classic example of transportation choice well-known from Domencich and McFadden (1975). Certainly we expect the value of time and money to be correlated across the population of commuters. But for a single commuter making a choice between car and bus on a specific morning, we do not require a specific relationship between the disutility of commuting time and the marginal utility of income she happens to "draw" from her random utility urn on that particular morning. This gives us some latitude when we choose a distribution for the unobserved parts of her utilities of various commuting alternatives.

We have much less of this latitude when we think of random preferences over lottery pairs. The spirit of random preferences is that every preference ordering drawn from the urn obeys all properties of the preference structure (Loomes and Sugden 1995). We demand, for instance, that

[18] The ratio relationship here is a generalization of the well-known fact that the ratio of independent chi-square variates follows an F distribution. Chi-square variates are gamma variates with common scale parameter $\kappa = 2$. In fact, a beta-prime variate can be transformed into an F variate: If x is a beta-prime variate with parameters a and b, then bx/a is an F variate with degrees of freedom $2a$ and $2b$. This is convenient because almost all statistics software packages contain thorough call routines for F variates, but not necessarily any call routines for beta-prime variates.

every vector of outcome utilities drawn from the urn respects monotonicity in z; this implies that the joint distribution of u_2^n and u_3^n must have the property that $u_3^n \geq u_2^n \geq 1$. Moreover, the assumptions we make about the v_c^n must be probabilistically consistent across pair contexts. Choosing a joint distribution of u_2^n and u_3^n immediately implies exact commitments regarding the distribution of any and all functions of u_2^n and u_3^n. The issue does not arise in a data set where subjects make choices from pairs on just one context, as in Loomes, Moffatt and Sugden (2002): In this simplest of cases, <u>any</u> distribution of v_c^n on $\mathbb{R}^+$, including the Lognormal choice they make, is a wholly legitimate hypothesis. But as soon as each subject makes choices from pairs on several different overlapping contexts, random preferences are much more exacting. Unless we can specify a joint distribution of g_1^n and g_2^n that implies it, we are <u>not</u> entitled (for instance) to assume that v_c^n follows Lognormal distributions in all of three overlapping contexts for a single subject.[19] Put differently, our choice of a joint distribution for v_{-2}^n and v_{-3}^n has inescapable implications for the distribution of v_{-0}^n. Carbone (1997) correctly saw this in her random preference treatment of the EU structure. Under these circumstances, a clever choice of the joint distribution of g_1^n and g_2^n is necessary.

The random preference model can be quite limiting in practical applications. For instance, notice that I have not specified a distribution of v_{-1}^n and hence have no method for applying equation (B.3) to pairs on the context (0,2,3). Fully a quarter of the data from experiments such as Hey and Orme (1994) are on that context. On the context −1 = (0,2,3), $v_{-1}^n = g_2^n/(1+g_1^n)$. As far as I am aware, the following is a true statement, though I may yet see it disproved.

[19] Although ratios of lognormal variates are lognormal, there is no similar simple parametric family for sums of lognormal variates. The independent gammas with common scale are the only workable choice I am aware of.

Conjecture: There is no nondegenerate joint distribution of g_1^n and g_2^n on $(\mathbb{R}^+)^2$ such that

g_1^n, $g_1^n + g_2^n$, $g_2^n/(g_1^n+1)$ and g_2^n/g_1^n all have tractable parametric distributions.

This is why I limit myself here to the three-fourths of these data sets that are choices from pairs on the contexts (0,1,2), (0,1,3) and (1,2,3): These are the contexts that the "independent gamma model" of random preferences developed above can be applied to. There are no similar practical modeling constraints on RDU strict, strong or contextual utility models (a considerable practical point in their favor); these models are easily applied to choices on any context. Note, however, that Loomes, Moffat and Sugden's (2002) technique could be sidestepped by (say) simulated maximum likelihood techniques.

Appendix C: Estimation—an EU and Strong Utility Illustration

Though the Hey and Orme (1994) experiment contains no FOSD pairs, Moffatt and Peters (2001) found evidence of nonzero tremble probabilities using it, so I include a tremble probability in all models. However, using Hey's (2001) still larger data set, I find no evidence that tremble probabilities vary across subjects. Therefore, I assume that $\omega^n = \omega$ for all subjects n, so that likelihood functions are built from the probabilities $\mathrm{P}_{mc}^n = (1-\omega)P_{mc}^n + \omega/2$.

Let (EU,Strong) denote an expected utility structure with the strong utility stochastic model in which $\psi^n = (u_2^n, u_3^n, \lambda^n, \omega)$ is subject n's true parameter vector governing her choices from pairs. Let $J(\psi \mid \theta)$ denote the joint c.d.f. governing the distribution of $\psi = (u_2, u_3, \lambda, \omega)$ in the sampled population, where θ are parameters governing the shape and location of J. Let θ^* be the true value of θ in that population. We want an estimate $\hat{\theta}$ of θ^*: This is random parameters estimation. We need a procedure for choosing a reasonable and tractable form for J that appears to characterize main features of the joint distribution of ψ in the sample. What follows illustrates this procedure for the (EU,Strong) specification: See Wilcox (2007a) for a more elaborated description of the procedure and exact details of all specifications estimated here.

Suppressing the subject index n, the (EU,Strong) specification is, at the individual level,

$$\mathrm{P}_{mc} = (1-\omega)\Lambda\big(\lambda[(s_{m1}-t_{m1})u_{1c} + (s_{m2}-t_{m2})u_{2c} + (s_{m3}-t_{m3})u_{3c}]\big) + \omega/2, \tag{C.1}$$

where $\Lambda(x) = [1+\exp(-x)]^{-1}$ is the Logistic c.d.f. (consistently employed as the function $H(x)$ for the strong, strict and contextual utility models). Equation C.1 introduces the notation $u_{ic} = u(z_{ic})$. In terms of the underlying utility parameters u_2 and u_3 of a subject, (u_{1c}, u_{2c}, u_{3c}) is $(1, u_2, u_3)$ for pairs on context $c = -0 = (1,2,3)$, $(0,1,u_3)$ for pairs on context $c = -2 = (0,1,3)$; and $(0,1,u_2)$ for pairs on context $c = -3 = (0,1,2)$.

Begin with individual estimation of a simplified version of (C.1), using sixty-eight of HO's eighty subjects.[20] This initial estimation gives an impression of the form of the joint distribution of ψ, and how $J(\psi \mid \theta)$ may be chosen to represent it. At this initial step, ω is not estimated, but is instead set equal to 0.04 in (C.1) for all subjects.[21] (Estimation of ω is undertaken later in the random parameters estimation.) The log likelihood function for subject *n*, in terms of P_{mc} in (C.1) with $\omega = 0.04$, is

$$LL^n(u_2,u_3,\lambda) = \sum_{mc} y^n_{mc} \ln(\mathrm{P}_{mc}) + (1-y^n_{mc})\ln(1-\mathrm{P}_{mc}) . \qquad \text{(C.2)}$$

Maximizing this in u_2, u_3 and λ yields initial estimates $\tilde{\psi}^n = (\tilde{u}^n_2, \tilde{u}^n_3, \tilde{\lambda}^n, 0.04)$ for each subject *n*. Figure 8 graphs $\ln(\tilde{u}^n_2 - 1)$, $\ln(\tilde{u}^n_3 - 1)$ and $\ln(\tilde{\lambda}^n)$ against their first principal component, which accounts for about 69 percent of their collective variance.[22] The figure also shows regression lines on the first principal component. The Pearson correlation between $\ln(\tilde{u}^n_2 - 1)$ and $\ln(\tilde{u}^n_3 - 1)$ is fairly high (0.848), and since these are estimates containing some pure sampling error, it appears that an assumption of perfect correlation between them in the underlying population may be roughly correct. Therefore, I make this assumption about the joint distribution of ψ in the population. While $\ln(\tilde{\lambda}^n)$ appears to share some variance with $\ln(\tilde{u}^n_2 - 1)$

[20] There are twelve subjects in the HO data with few or no choices of the riskier lottery in any pair. They can ultimately be included in random parameters estimations, but at this initial stage of individual estimation it is either not useful (due to poor identification) or simply not possible to estimate models for these subjects.
[21] Estimation of ω is a nuisance at the individual level. Trembles are rare enough that individual estimates of ω are typically zero for individuals. Even when estimates are nonzero, the addition of an extra parameter to estimate increases the noisiness of the remaining estimates and hides the pattern of variance and covariance of these parameters that we wish to see at this step.
[22] Two hugely obvious outliers have been removed for both the principal components extraction and the graph.

and $\ln(\tilde{u}_3^n - 1)$ (Pearson correlations of −0.22 and −0.45, respectively), it obviously either has independent variance of its own or is estimated with relatively low precision.

These observations suggest that the joint distribution $J(\psi \mid \theta)$ of $\psi = (u_2, u_3, \lambda, \omega)$ can be characterized as generated by two independent standard normal deviates x_u and x_λ, as follows:

$u_2(x_u, \theta) = 1 + \exp(a_2 + b_2 x_u)$, $u_3(x_u, \theta) = 1 + \exp(a_3 + b_3 x_u)$,

$\lambda(x_u, x_\lambda, \theta) = \exp(a_\lambda + b_\lambda x_u + c_\lambda x_\lambda)$ and ω a constant,

where $(\theta, \omega) = (a_2, b_2, a_3, b_3, a_\lambda, b_\lambda, c_\lambda, \omega)$ are parameters to be estimated. (C.3)

Then the (EU,Strong) model, conditional on x_u, x_λ and (θ, ω), becomes

$\mathrm{P}_{mc}(x_u, x_\lambda, \theta, \omega) =$

$(1-\omega)\Lambda\big(\lambda(x_u, x_\lambda, \theta)[(s_{m1} - t_{m1})u_{1c} + (s_{m2} - t_{m2})u_{2c} + (s_{m3} - t_{m3})u_{3c}]\big) + \omega/2$, where

$u_{1c} = 1$ if $c = -0$, $u_{1c} = 0$ otherwise, $u_{2c} = u_2(x_u, \theta)$ if $c = -0$, $u_{2c} = 1$ otherwise, and

$u_{3c} = u_2(x_u, \theta)$ if $c = -3$, $u_{3c} = u_3(x_u, \theta)$ otherwise. (C.4)

This implies the following random parameters log likelihood function in (θ, ω):

$$LL(\theta, \omega) = \sum\nolimits_n \ln\Big(\iint \Big(\prod\nolimits_{mc} \mathrm{P}_{mc}(x_u, x_\lambda, \theta, \omega)^{y_{mc}^n}[1 - \mathrm{P}_{mc}(x_u, x_\lambda, \theta, \omega)]^{1-y_{mc}^n}\Big) d\Phi(x_u)\, d\Phi(x_\lambda)\Big), \qquad \text{(C.5)}$$

where Φ is the standard normal c.d.f. and $\mathrm{P}_{mc}(x_u, x_\lambda, \theta, \omega)$ is as shown in (C.4).[23] The regression lines in Figure 8 provide starting values for maximizing (C.5). That is, initial estimates of the a and b coefficients in θ are the intercepts and slopes from the linear regressions of $\ln(\tilde{u}_2^n - 1)$, $\ln(\tilde{u}_3^n - 1)$ and $\ln(\tilde{\lambda}^n)$ on their first principal component; and the root mean squared error of the regression of $\ln(\tilde{\lambda}^n)$ on the first principal component provides an initial estimate of c_λ.

[23] Such integrations must be performed numerically in some manner for estimation. I use gauss-hermite quadratures, which are practical up to two or three integrals; for integrals of higher dimension, simulated maximum likelihood is usually more practical. Judd (1998) and Train (2003) are good sources for these methods.

Table 3 shows the results of maximizing (C.5) in (θ,ω). These estimates produce the log likelihood in the first column of the top row of Table 1. Note that wherever $\hat{b}_2 \neq \hat{b}_3$, very large (or small) values of the underlying standard normal deviate x_u imply a violation of monotonicity (that is $u_2 > u_3$). Rather than imposing $b_2 = b_3$ as a constraint on the estimations, I impose the weaker constraint $|(a_2 - a_3)/(b_3 - b_2)| > 4.2649$, making the estimated population fraction of such violations no larger than 10^{-5}. This constraint does not bind for the estimates shown in Table 3. Generally, it rarely binds, and is never close to significantly binding, for any of the strong, strict or contextual utility estimations done here.

Recall that the nonparametric treatment of utility avoids a fixed risk attitude across the outcome vector (0,1,2,3), as would be implied by a parametric form such as CARA or CRRA utility. The estimates shown in Table 3 imply a population in which about sixty-eight percent of subjects have a weakly concave utility function, with the remaining thirty-two percent have an inflected "concave then convex" utility function, closely resembling Hey and Orme's (1994) individual estimation results. That is, the random parameters estimation used here produces utility function heterogeneity much like that suggested by individual estimation.

A very similar procedure was used to select and estimate random parameters characterizations of heterogeneity for all models. As with the detailed example of the (EU,Strong) specification, all specifications with utility parameters u_2^n and u_3^n (strong, strict or contextual utility specifications) yield quite high Pearson correlations between $\ln(\tilde{u}_2^n - 1)$ and $\ln(\tilde{u}_3^n - 1)$ across subjects, and heavy loadings of these on first principal components of estimated parameter vectors $\tilde{\psi}^n$. Therefore, the population distributions of $\psi = (u_2, u_3, \lambda, \gamma, \omega)$ (strong, strict and contextual utility models, with $\gamma \equiv 1$ for EU) are in all cases modeled as having a perfect correlation between $\ln(u_2^n - 1)$ and $\ln(u_3^n - 1)$, generated by an underlying standard normal deviate x_u.

Similarly, individual estimations of random preference models where $\psi = (\phi_1, \phi_2, \kappa, \gamma, \omega)$ ($\gamma \equiv 1$ for EU) yield high Pearson correlations between $\ln(\tilde{\phi}_1^n)$ and $\ln(\tilde{\phi}_2^n)$ across subjects, and heavy loadings of these on first principal components of estimated parameter vectors $\tilde{\psi}^n$. So joint distributions of $\psi = (\phi_1, \phi_2, \kappa, \gamma, \omega)$ are assumed to have a perfect correlation between $\ln(\phi_1^n)$ and $\ln(\phi_2^n)$ in the population, generated by an underlying standard normal deviate x_ϕ.

In all cases, all other model parameters are characterized as possibly partaking of some of the variance represented by a normally distributed first principle component x_u (in strong, strict or contextual utility specifications) or x_ϕ (in random preference specifications), but also having independent variance represented by an independent standard normal variate, as with the example of λ in the (EU,Strong) specification as shown in (C.3).

For EU specifications, a likelihood function like (C.5) is maximized. RDU specifications add a third integration since these models allow for independent variance in γ (the Prelec weighting function parameter) through the addition of a third standard normal variate x_γ. Integrations are carried out by gauss-hermite quadrature. In all cases, starting values for these numerical maximizations are computed in the manner described for the (EU,Strong) model: Parameters in $\tilde{\psi}^n$ are regressed on their first principle component, and the intercepts and slopes of these regressions are the starting values for the *a* and *b* coefficients in the models, while the root mean squared errors of these regressions are the starting values for the *c* coefficients found in the equations for λ, κ and/or γ.

Acknowledgments

John Hey generously made his remarkable data sets available to me. I also thank John Hey and Pavlo Blavatskyy, Soo Hong Chew, Edi Karni, Ondrej Rydval, Peter Wakker, two anonymous referees and especially Glenn Harrison for conversation, commentary and/or help, though any errors here are solely my own. I thank the National Science Foundation for support under grant SES 0350565.

References

Aitchison, J., 1963, Inverse distributions and independent gamma-distributed products of random variables. Biometrika 50, 505-508.

Andersen, S., G. W. Harrison, M. I. Lau and E. E. Rutström, 2007, Eliciting risk and time preferences. Forthcoming in Econometrica.

Ballinger, T. P., and N. Wilcox, 1997, Decisions, error and heterogeneity. Economic Journal 107, 1090-1105.

Becker, G. M., M. H. DeGroot and J. Marschak, 1963a, Stochastic models of choice behavior. Behavioral Science 8, 41–55.

Becker, G. M., M. H. DeGroot and J. Marschak, 1963b, An experimental study of some stochastic models for wagers. Behavioral Science 8, 199-202.

Blavatskyy, P. R., 2007, Stochastic expected utility theory. Journal of Risk and Uncertainty 34, 259-286.

Block, H. D. and J. Marschak, 1960, Random orderings and stochastic theories of responses, in I. Olkin et al., (Eds.), Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford University Press, Stanford, pp. 97-132.

Buschena, D. E. and D. Zilberman, 2000, Generalized expected utility, heteroscedastic error, and path dependence in risky choice. Journal of Risk and Uncertainty 20, 67-88.

Busemeyer, J. and J. Townsend, 1993, Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review 100, 432-59.

Busemeyer, J. and Y-M. Wang, 2000, Model comparisons and model selections based on generalization criterion methodology. Journal of Mathematical Psychology 44, 171-89.

Camerer, C., 1989, An experimental test of several generalized expected utility theories. Journal of Risk and Uncertainty 2, 61-104.

Camerer, C. and T-H. Ho, 1999, Experience weighted attraction learning in normal-form games. Econometrica 67:827-74.

Carbone, E., 1997, Investigation of stochastic preference theory using experimental data. Economics Letters 57, 305-311.

Chew, S. H., 1983, A generalization of the quasilinear mean with applications to the measurement of income inequality and decision theory resolving the Allais paradox. Econometrica 51, 1065-1092.

Chew, S. H., E. Karni and Z. Safra, 1987, Risk aversion in the theory of expected utility with rank-dependent preferences. Journal of Economic Theory 42, 370-81.

Debreu, G., 1958, Stochastic choice and cardinal utility. Econometrica 26, 440-444.

Domencich, T., and D. McFadden, 1975, Urban travel demand: A behavioral analysis. Amsterdam: North-Holland.

Edwards, W., 1954, A theory of decision making. Psychological Bulletin 51, 380-417.

Gravetter, F. and G. R. Lockhead, 1973, Criterial range as a frame of reference for stimulus judgment. Psychological Review 80, 203-216.

Gul, F., and W. Pesendorfer, 2006, Random expected utility. Econometrica 74, 121-146.

Halff, H. M., 1976, Choice theories for differentially comparable alternatives. Journal of Mathematical Psychology 14, 244-246.

Harrison, G. W. and E. E. Rutström, 2005, Expected utility theory and prospect theory: One wedding and a decent funeral. Working paper 05-18, Department of Economics, College of Business Administration, University of Central Florida.

Hartman, E. B., 1954, The influence of practice and pitch-distance between tones on the absolute identification of pitch. American Journal of Psychology 67, 1-14.

Hey, J. D., 1995, Experimental investigations of errors in decision making under risk. European Economic Review 39, 633-640.

Hey, J. D., 2001, Does repetition improve consistency? Experimental Economics 4 5-54.

Hey, J. D. and E. Carbone, 1995, Stochastic choice with deterministic preferences: An experimental investigation. Economics Letters 47, 161–167.

Hey, J. D. and C. Orme, 1994, Investigating parsimonious generalizations of expected utility theory using experimental data. Econometrica 62, 1291-1329.

Hilton, R. W. 1989. Risk attitude under random utility. Journal of Mathematical Psychology 33, 206-222.

Holt, C. A. and S. K. Laury, 2002, Risk aversion and incentive effects. American Economic Review 92, 1644-1655.

Hutchinson, T. P. and C. D. Lai, 1990, Continuous bivariate distributions, emphasizing applications. Rumsby Scientific Publishers, Adelaide, Australia.

Judd, K. L., 1998, Numerical methods in economics. MIT Press, Cambridge U.S.A.

Loomes, G., 2005, Modeling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. Experimental Economics 8, 301-323.

Loomes, G. and R. Sugden, 1995, Incorporating a stochastic element into decision theories. European Economic Review 39, 641-648.

Loomes, G. and R. Sugden, 1998, Testing different stochastic specifications of risky choice. Economica 65, 581-598.

Loomes, G., P. Moffatt and R. Sugden, 2002, A microeconometric test of alternative stochastic theories of risky choice. Journal of Risk and Uncertainty 24, 103-130.

Luce, R. D. and P. Suppes, 1965, Preference, utility and subjective probability, in R. D. Luce, R. R. Bush and E. Galanter, (Eds.), Handbook of mathematical psychology Vol. III. Wiley, New York, pp. 249-410.

Machina, M., 1985, Stochastic choice functions generated from deterministic preferences over lotteries. Economic Journal 95, 575-594.

McKelvey, R. and T. Palfrey, 1995, Quantal response equilibria for normal form games. Games and Economic Behavior 10, 6-38.

Moffatt, P., 2005, Stochastic choice and the allocation of cognitive effort. Experimental Economics 8, 369-388.

Moffatt, P. and S. Peters, 2001, Testing for the presence of a tremble in economics experiments. Experimental Economics 4, 221-228.

Myers, J. L. and E. Sadler, 1960, Effects of range of payoffs as a variable in risk taking. Journal of Experimental Psychology 60, 306-309.

Papke, L. E., and J. M. Wooldridge, 1996, Econometric methods for fractional response variables with an application to 401(k) plan participation rates. Journal of Applied Econometrics 11, 619-632.

Parducci, A., 1965, Category judgment: A range-frequency model. Psychological Review 72, 407-418.

Parducci, A., 1974, Contextual effects: A range-frequency analysis, in E. C. Carterette and M. P. Friedman, (Eds.), Handbook of perception Vol. 2. Academic Press, New York, 127-141.

Pollack, I., 1953, The information of elementary auditory displays, II. Journal of the Accoustical Society of America 25, 765-769.

Pratt, J. W., 1964, Risk aversion in the small and in the large. Econometrica 32, 122-136.

Prelec, D., 1998, The probability weighting function. Econometrica 66, 497-527.

Quiggin, J, 1982, A theory of anticipated utility. Journal of Economic Behavior and Organization 3, 323-343.

Rothschild, M. and J. E. Stiglitz, 1970, Increasing risk I: A definition. Journal of Economic Theory 2, 225-243.

Saha, A, 1993, Expo-Power Utility: A 'Flexible' Form for Absolute and Relative Risk Aversion. American Journal of Agricultural Economics 75, 905-913.

Starmer, C. and R. Sugden, 1989, Probability and juxtaposition effects: An experimental investigation of the common ratio effect. Journal of Risk and Uncertainty 2, 159-78.

Tversky, A. and J. E. Russo, 1969, Substitutability and similarity in binary choices. Journal of Mathematical Psychology 6, 1-12.

Tversky, A. and D. Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

Vuong, Q., 1989, Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Wilcox, N., 2007a, Stochastic models for binary discrete choice under risk: A critical stochastic modeling primer and econometric comparison. Forthcoming in J.C. Cox and G. W. Harrison, (Eds.), Research in experimental economics Vol. 12: Risk aversion in experiments. JAI Press, Greenwich, CT.

Wilcox, N., 2007b, Predicting risky choices out-of-context: A Monte Carlo study. Working Paper, University of Houston Department of Economics.

Wooldridge, J. M., 2002, Econometric analysis of cross section and panel data. MIT Press, Cambridge U.S.A.

Table 1. Log likelihoods of random parameters characterizations of the models in the Hey and Orme sample

		Estimated on all three contexts	Estimated on contexts (0,1,2) and (0,1,3)
Structure	Stochastic Model	Log Likelihood on all three contexts (in-sample fit)	Log likelihood on context (1,2,3) (out-of-sample fit)
EU	Strong Utility	−5311.44	−2409.38
	Strict Utility	−5448.50	−2373.12
	Contextual Utility	−5297.08	−2302.55
	Random Preferences	−5348.36	−2356.60
RDU	Strong Utility	−5207.81	−2394.75
	Strict Utility	−5306.48	−2450.41
	Contextual Utility	−5190.43	−2281.36
	Random Preferences	−5218.00	−2335.55

Table 2. Vuong (1989) non-nested tests between model pairs,
by structure, stochastic model and in-sample versus out-of-sample fit.

EU Structure							
Estimated on all three contexts, and comparing fit on all three contexts (in-sample fit comparison)				Estimated on contexts (0,1,2) and (0,1,3), and comparing fit on context (1,2,3) (out-of-sample fit comparison)			
	Random Prefs.	Strong Utility	Strict Utility		Random Prefs.	Strong Utility	Strict Utility
Contextual Utility	$z = 1.723$ $p = 0.042$	$z = 0.703$ $p = 0.241$	$z = 6.067$ $p < 0.0001$	Contextual Utility	$z = 4.387$ $p < 0.0001$	$z = 3.044$ $p = 0.0012$	$z = 2.739$ $p = 0.0031$
Random Prefs.	—	$z = -1.574$ $p = 0.058$	$z = 5.419$ $p < 0.0001$	Random Prefs.	—	$z = 1.639$ $p = 0.051$	$z = 1.422$ $p = 0.078$
Strong Utility	—	—	$z = 5.961$ $p < 0.0001$	Strong Utility	—	—	$z = 0.028$ $p = 0.49$
RDU Structure							
Estimated on all three contexts, and comparing fit on all three contexts (in-sample fit comparison)				Estimated on contexts (0,1,2) and (0,1,3), and comparing fit on context (1,2,3) (out-of-sample fit comparison)			
	Random Prefs.	Strong Utility	Strict Utility		Random Prefs.	Strong Utility	Strict Utility
Contextual Utility	$z = 0.981$ $p = 0.163$	$z = 0.877$ $p = 0.190$	$z = 4.352$ $p < 0.0001$	Contextual Utility	$z = 3.879$ $p < 0.0001$	$z = 3.304$ $p = 0.0005$	$z = 5.978$ $p < 0.0001$
Random Prefs.	—	$z = -0.44$ $p = 0.330$	$z = 3.808$ $p < 0.0001$	Random Prefs.	—	$z = 1.652$ $p = 0.049$	$z = 3.831$ $p < 0.0001$
Strong Utility	—	—	$z = 5.973$ $p < 0.0001$	Strong Utility	—	—	$z = 3.918$ $p < 0.0001$

Notes: Positive z means the row stochastic model fits better than the column stochastic model.

Table 3. Random parameters estimates of the (EU,Strong) model, using choice data from the contexts (0,1,2), (0,1,3) and (1,2,3) of the Hey and Orme (1994) sample.

Structural and stochastic parameter models	Distributional parameter	Initial estimate	Final estimate	Asymptotic standard error	Asymptotic t-statistic
$u_2 = 1 + \exp(a_2 + b_2 x_u)$	a_2	−1.2	−1.28	0.0411	−31.0
	b_2	0.57	0.514	0.0311	16.5
$u_3 = 1 + \exp(a_3 + b_3 x_u)$	a_3	−0.51	−0.653	0.0329	−16.9
	b_3	0.63	0.657	0.0316	20.8
$\lambda = \exp(a_\lambda + b_\lambda x_u + c_\lambda x_\lambda)$	a_λ	3.2	3.39	0.101	33.8
	b_λ	−0.49	−0.658	0.124	−5.32
	c_λ	0.66	0.584	0.0571	10.2
ω constant	ω	0.04	0.0446	0.0105	4.26
Log likelihood = −5311.44					

Notes: x_u and x_λ are independent standard normal variates. Standard errors are calculated using the "sandwich estimator" (Wooldridge 2002) and treating all of each subject's choices as a single "super-observation," that is, using degrees of freedom equal to the number of subjects rather than the number of subjects times the number of choices made.

Figure 1. Articles with text references to risk aversion and substitutability relations, 1977-2001 (JSTOR Economics journals and selected Business journals) and Citations of Pratt (1964), 1977-2006 (Science and Social Science Citation Indices).

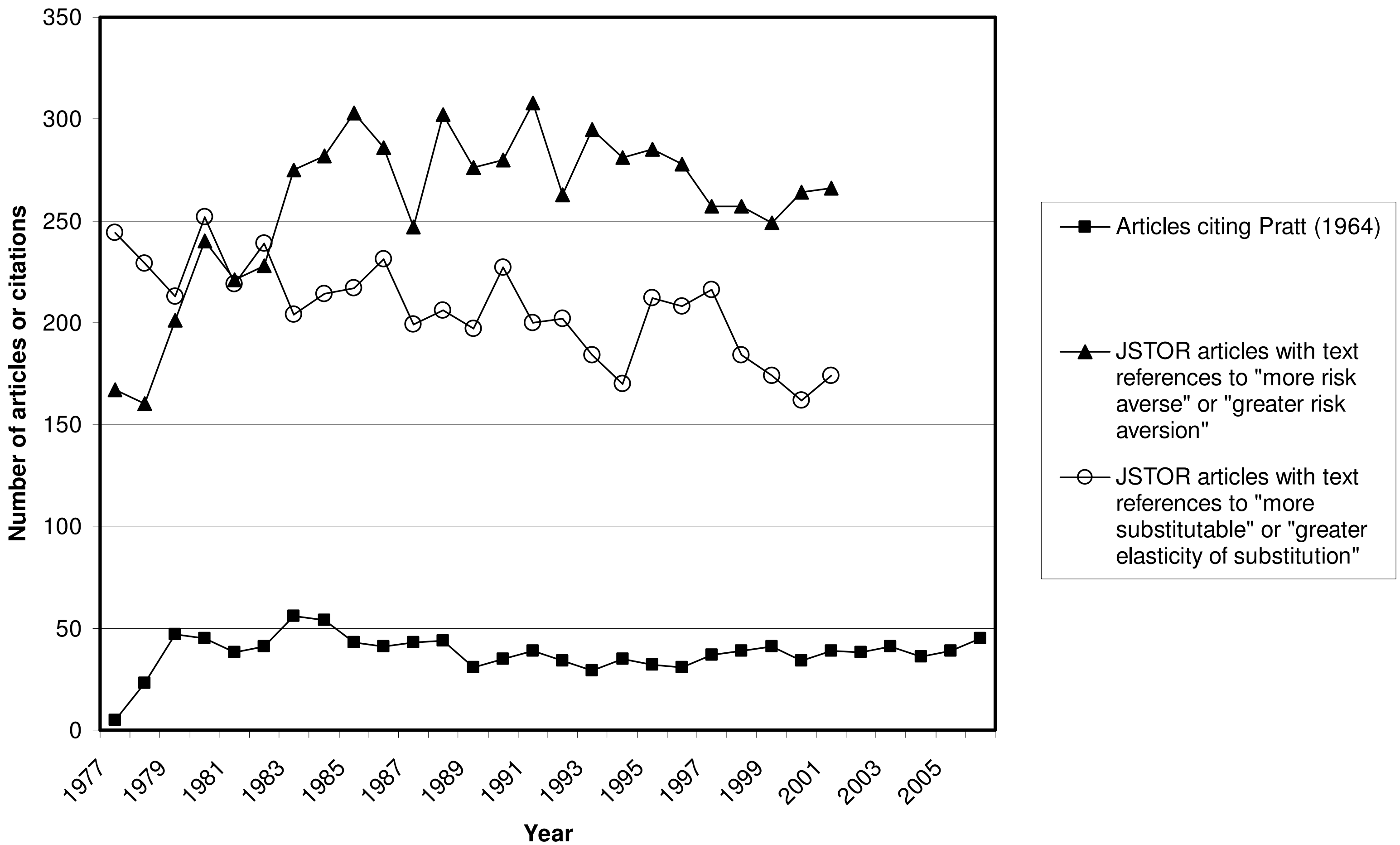

Figure 2. Behavior of five CRRA EU V-differences using various transformations, with homoscedastic precision: MPS pair 1 on context (0,50,100) from Hey (2001)

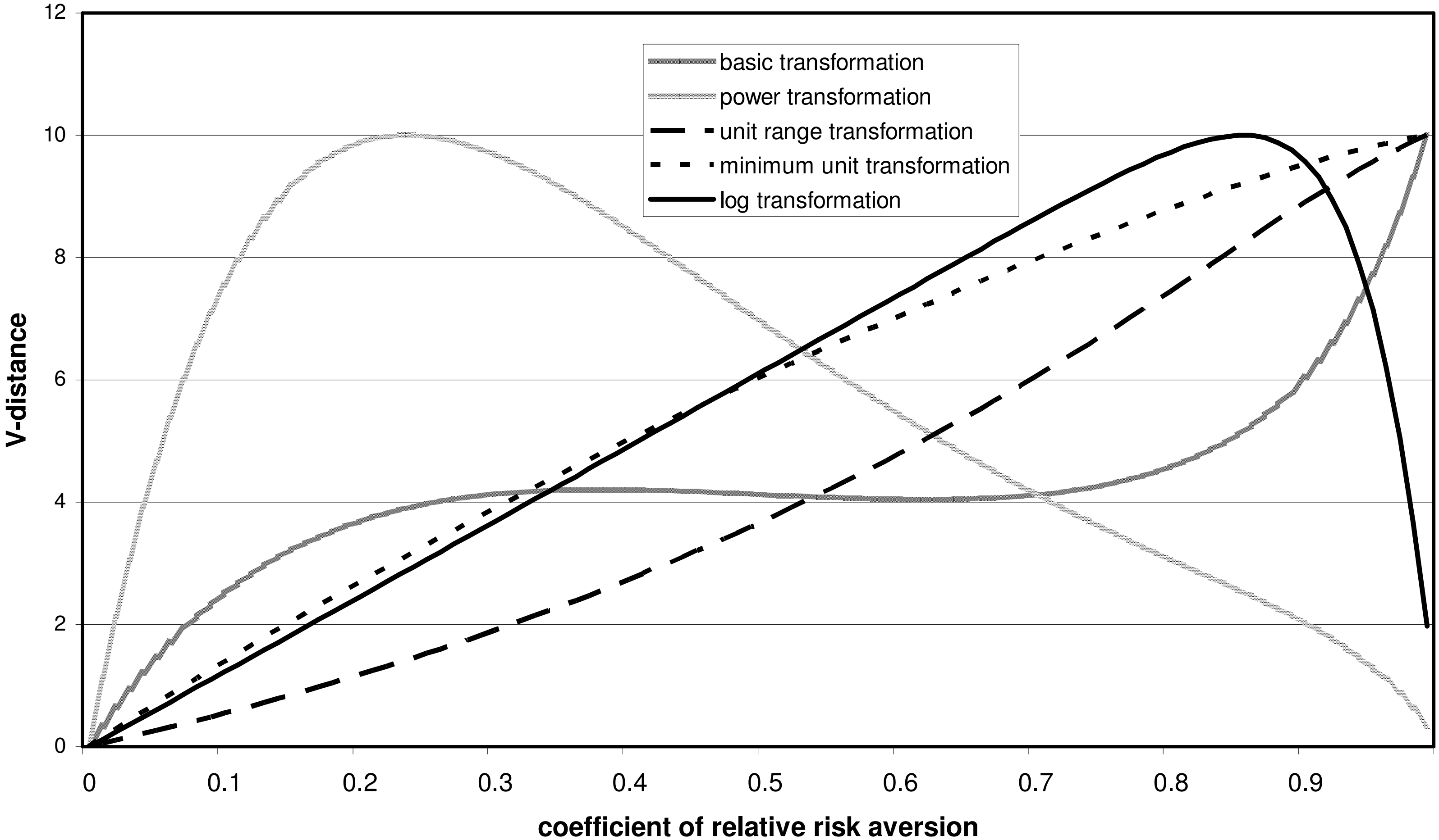

Figure 3. Behavior of five CRRA EU V-differences using various transformations, with homoscedastic precision: MPS pair 1 on context (50,100,150) from Hey (2001)

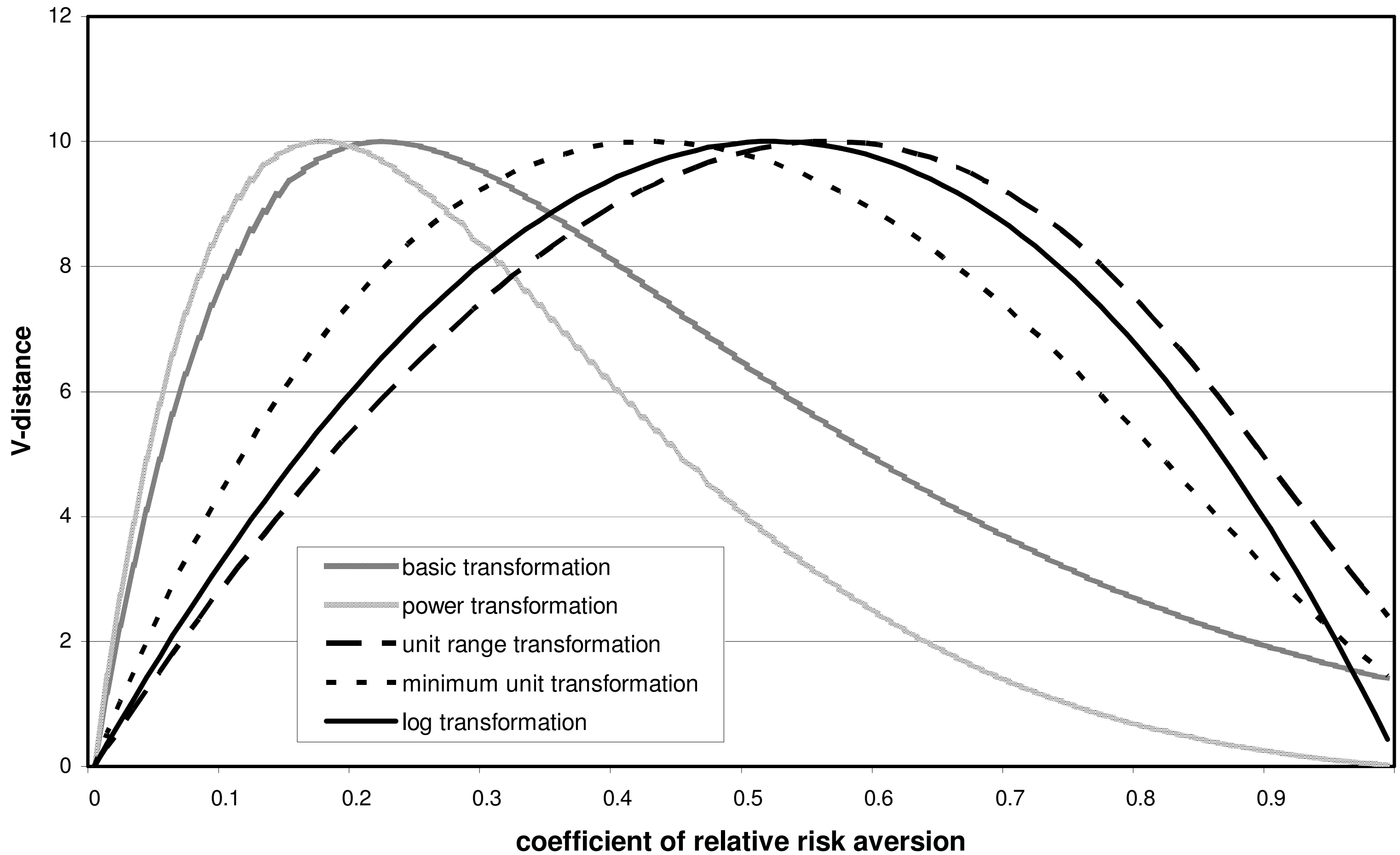

Figure 4. How agent heteroscedasticity might allow MRA to imply SMRA

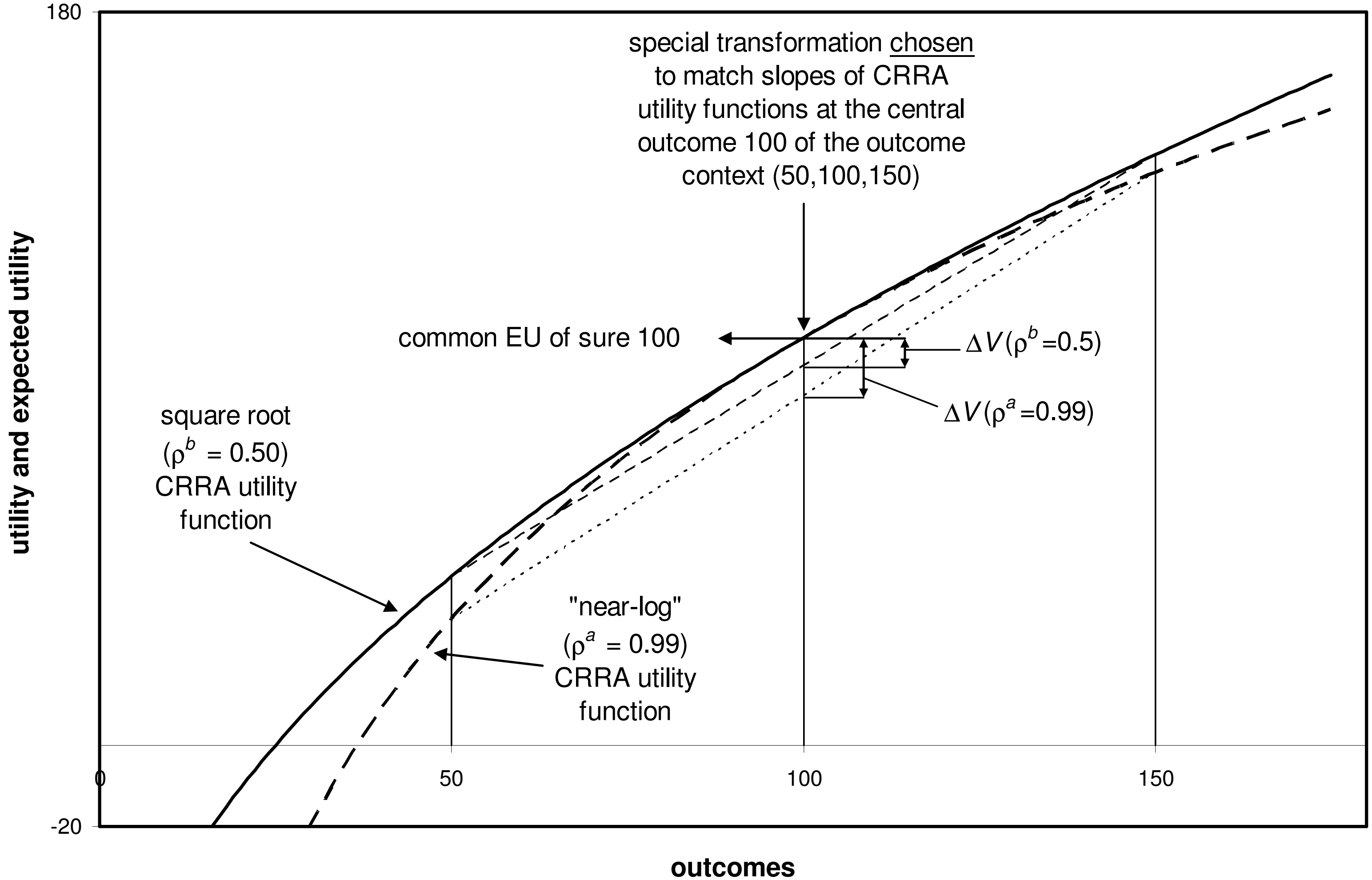

Figure 5. Relationship between risk aversion and precision in Hey (2001), CRRA EU estimation.

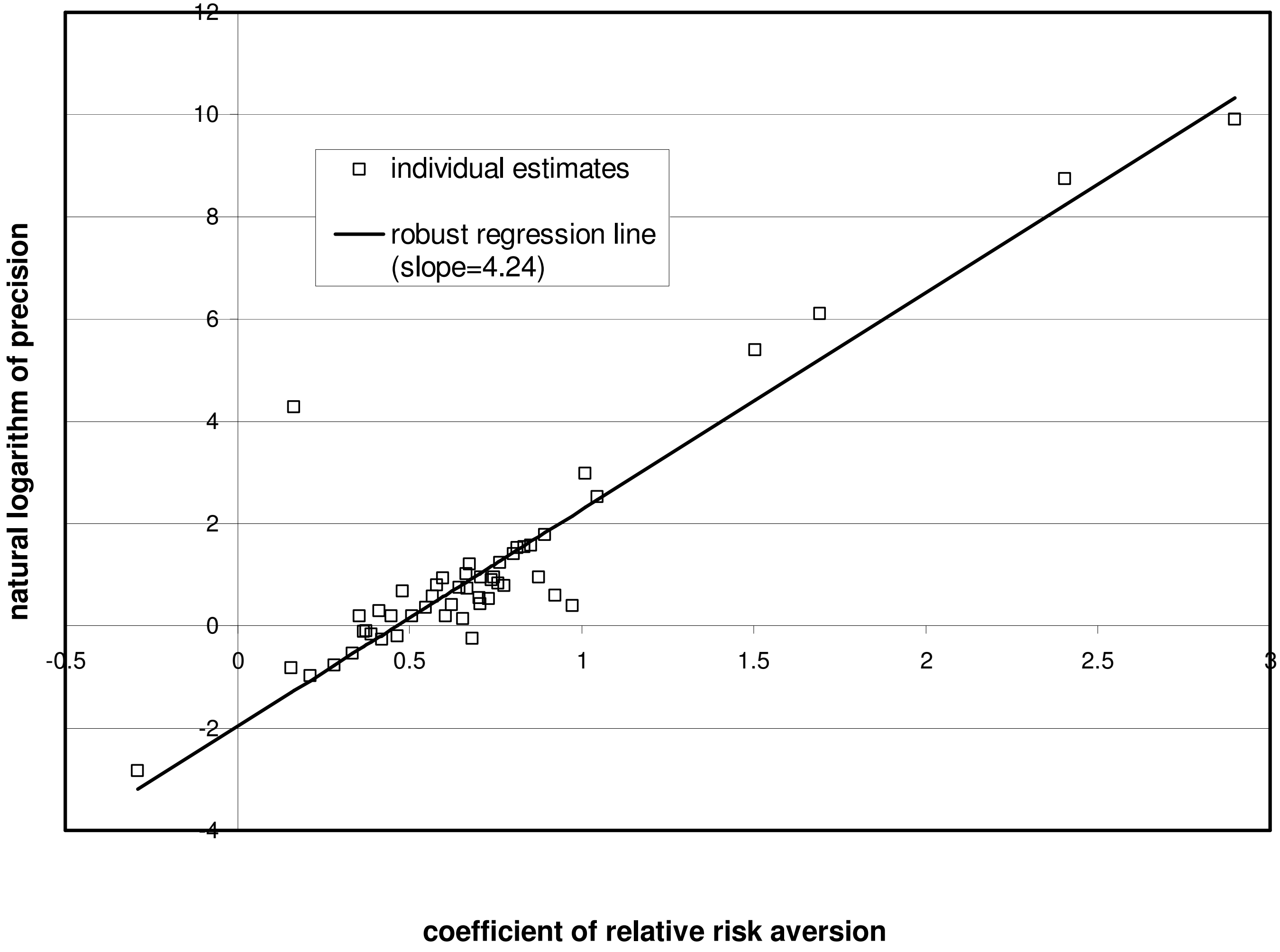

Figure 6: SMRA-compatible CRRA Utility Functions, slope and height matched at *z* = 69.41 = *exp*(4.24) (point estimate of α = 4.24)

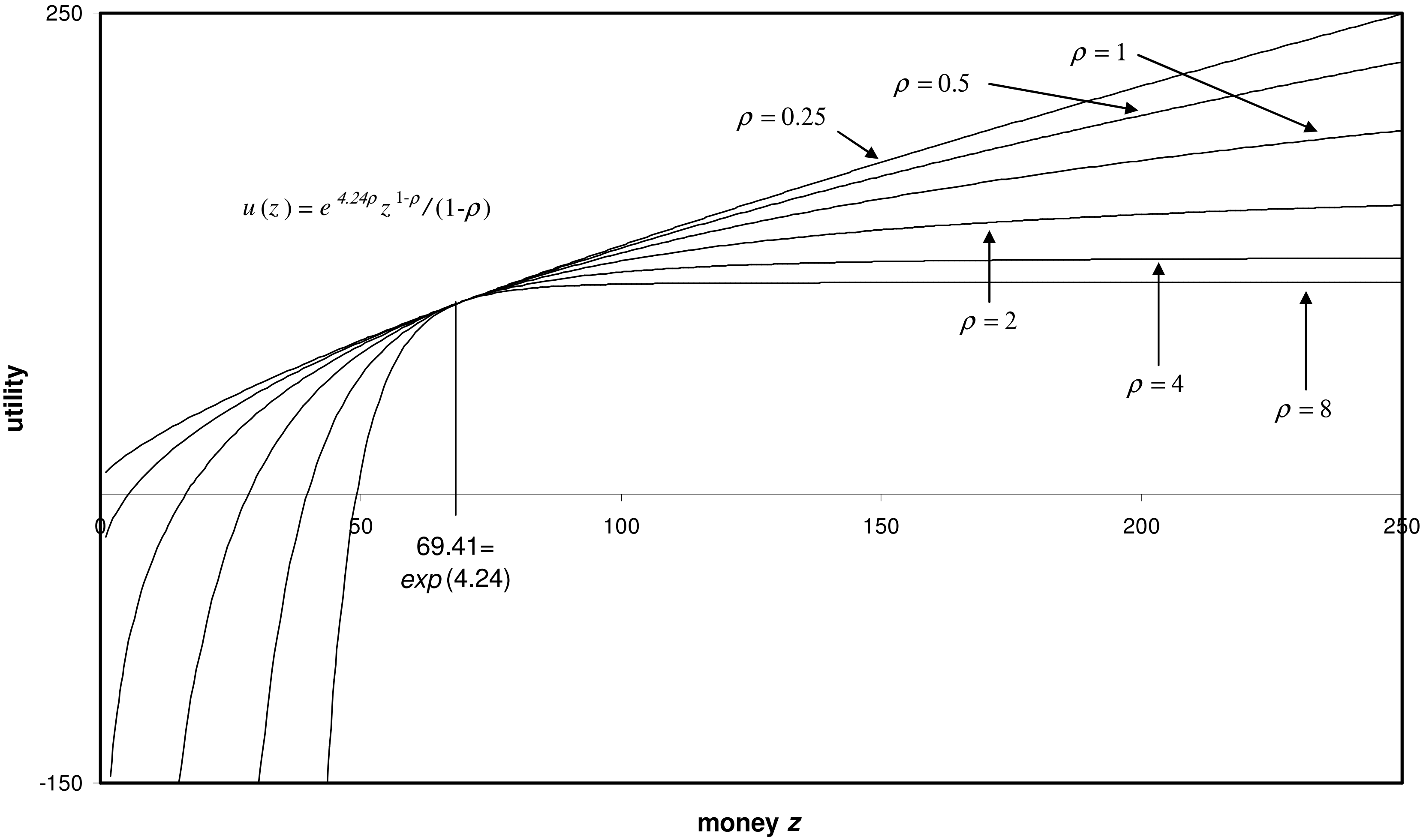

Figure 7: Behavior of CRRA EU V-Differences using SMRA-Compatible transformation: MPS pair 1 on the new context (100,150,200), at various values of α.

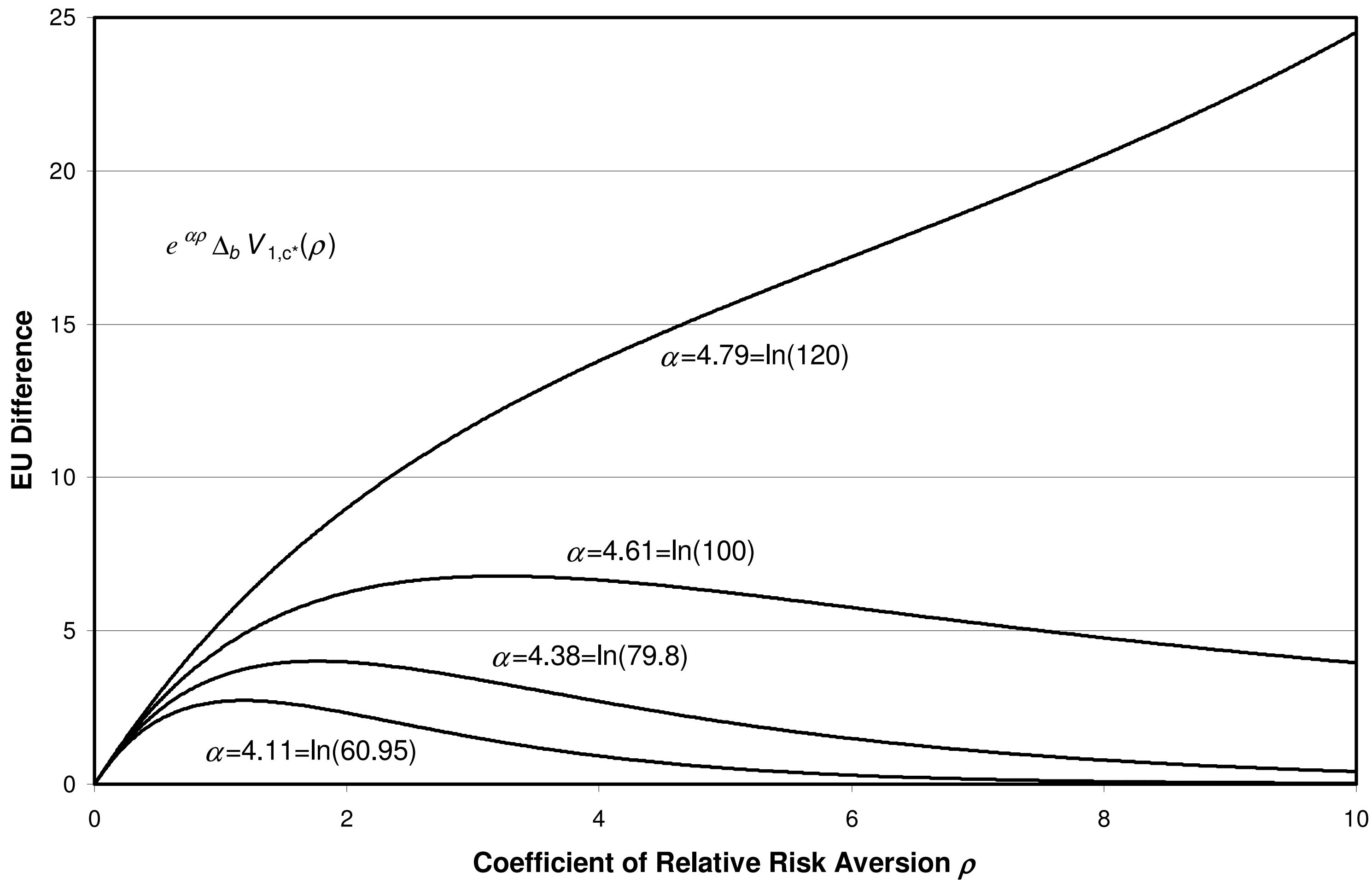

Figure 8. Shared variance of initial individual parameter estimates using the (EU,Strong) specification