Exp Econ (2007) 10:123–141
DOI 10.1007/s10683-006-9136-y

Stability of risk preference parameter estimates within the Becker-DeGroot-Marschak procedure

Duncan James

Received: 18 October 2004 / Revised: 7 February 2006 /
Accepted: 1 March 2006 / Published online: 15 February 2007

Abstract This paper reports new data from both selling and buying versions of the Becker-DeGroot-Marschak (BDM) procedure. First, when using the selling version of BDM, the cross-sectional mean of CRRA risk preference parameter estimates shifts from a value consistent with "as if" risk-seeking behavior in the early baseline to a value closer to "as if" risk neutrality in the late baseline. Second, when using the buying version of BDM, the cross-sectional mean of CRRA risk preference parameter estimates does not appear to change over time in a statistically significant manner. The cross-sectional mean from the late baseline of the buying version of BDM is closer to "as if" risk neutrality and to the late baseline estimates from the selling version of BDM than it is to either early baseline estimates from the selling version of BDM or typical estimates from the first price auction. Use of dominated offers is correlated with deviations from "as if" risk neutrality; this suggests the possibility that the early deviations from "as if" risk neutrality reflect errors.

Keywords Risk · Expected utility · Learning

JEL Classification D80

1 Introduction

There is a growing body of empirical results that document risk preference parameter inconsistency across institutions. Clear instances of the same subjects appearing "as if" risk averse in one institution (e.g., first price sealed bid auction) and "as if" risk seeking in another institution (e.g., Becker-DeGroot-Marschak or the English oral

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s10683-006-9136-y.

D. James (✉)
Fordham University, Department of Economics, Bronx, NY 10458, USA

auction) have been documented by Isaac and James (2000a) and Berg et al. (2005). This presents a puzzle in the interpretation of experimental data that differ systematically from risk neutral theoretical predictions. That is, if data from a particular institution cluster on one side of the risk neutral prediction for that institution, does that tell us about subjects' risk preferences? What if the same subjects are found to cluster on one side of the risk neutral prediction in a different institution—but that clustering is on the *other* side of risk neutrality?

One way to attempt to shed further light on this issue is to gather more data, under different circumstances, than were available when the puzzle was first noticed. A possible approach to this is to concentrate on a particular institution, and to measure possible influences on the data from that institution other than the subjects' own "as if" risk preference parameters.[1]

For instance, might confusion on the part of the subjects lead to systematic errors that fall asymmetrically relative to the appropriate risk neutral prediction, but that are not necessarily indicative of the subjects' risk preferences? Confusion in this sense might be hard to define or to measure, but its obverse—learning—might perhaps be documented without undue reliance on compound hypotheses about either the precise type of confusion which existed initially or the precise process by which learning occurs.

The approach employed by this paper is then to document the distribution and stability of, and possible influences on, risk preference parameter estimates based on relatively long time series of experimental data from the Becker-DeGroot-Marschak procedure. A key benefit to employing long time series from a single institution is that it allows for learning to take place, so that the effects of learning might be measured. The effects of learning are in turn measured by a simple and minimal standard of rationality: adherence to a dominant strategy.

The experiments in this paper have relevance for several different streams of research. While this paper is primarily related to the literature on risk preference parameter estimation, there are also implications for research on learning to play dominant strategies, and for research on the effect of payoff transformations on risky choice behavior. The rest of this paper is as follows. Section 2 reviews the related literature. Section 3 outlines the research questions of interest. Section 4 discusses the experimental design. Section 5 presents the results. Section 6 concludes.

2 Background

What can risk preference parameter estimates from experimental data tell us? Different researchers would nominate different answers to that question, based on different approaches to model building.

Some might suggest that such risk preference parameter estimates tell us about individuals' biologically innate risk preferences. As an example of this, Harlow and

[1] The phrase "as if" is used to acknowledge that this work deals with risk preferences as estimates of parameters which are in turn modeling constructs, and that this work does not necessarily claim that, say, a particular parameterization of expected utility is the only possible representation of risky choice. With readers' understanding, the phrase "as if" will be omitted from the rest of the paper.

Brown (1990) entertain the possibility of a correlation between results from economics experiments and biomedical measurements. In contrast to this, there is a long-standing tradition in economics that theoretical concepts such as risk preferences are regarded as "as if" modeling constructs, to be judged by their predictive value only (Friedman, 1953).

Another point of contention in this area is the role of income versus wealth in defining and measuring risk preferences. Rabin (2000) casts doubt upon inferences made about utility functions defined over *wealth* when those inferences are based on data from economics experiments using "low" monetary stakes. Specifically, Rabin demonstrates that the estimates derived in experimental settings imply absurdly risk averse behavior if applied to (higher stakes) field situations. In counterpoint to this, Cox and Sadiraj (2006) examine the use of utility functions defined over *income* rather than wealth as being not subject to Rabin's critique.

However, findings of cross-institutional parameter inconsistency (such as those of Isaac and James or Berg, Dickhaut, and McCabe) bring up a puzzle that is fundamental in nature. That is because one point upon which *all* parties to the various discussions on risk preferences would agree is that risk preferences for an individual are typically not thought to be a function of the institution in which the individual operates. Thus whether one believes that risk preferences actually exist, or are an "as if" modeling fiction, or whether one believes that deviations from risk neutral behavior due to risk preferences can be accurately measured in a typical experimental setting, or not, one still needs to make further inquiry with regard to the asymmetric clustering of data (relative to risk neutrality) in each institution and the possible causes thereof.

As one example of such asymmetry, take the experimental research program surrounding the Constant Relative Risk Aversion Model of bidding in the first price auction (hereafter, CRRAM). Cox, Smith, and Walker (1988) use the CRRAM model to estimate values for subjects' risk-preference parameter, r_i, and find that r_i is less than one for over 90 percent of their subjects (statistically significantly less than one for 70 percent of their subjects). An r_i less than one implies risk aversion if the subject has CRRA preferences. Other sets of experiments by those researchers, and by other researchers including Schorvitz (1998), Isaac and James (2000a), Dorsey and Razzolini (2003), and Berg et al. (2005) have replicated the Cox, Smith, and Walker results. While these results can be argued not to demonstrate risk aversion, even then it remains an interesting question why the estimates take on the asymmetric pattern replicated across these studies.

A contrasting example of asymmetry in estimates of r_i involves the Becker-DeGroot-Marschak procedure (hereafter, BDM). BDM places the subject in a second-price auction where the other participant is not another person, but a draw from a uniform distribution, and where the object being auctioned is a lottery. Because it is a dominant strategy to reveal one's true valuation in a second-price auction, the subject should reveal what the lottery is truly worth to her. Given this observed certainty equivalent and the experimenter-controlled design parameters, one could solve for (estimate) the unobserved CRRA r_i parameter of the subject. Using BDM, risk preference parameter estimates greater than one—that is, consistent with *risk seeking*—have been observed (Harrison, 1990; Isaac and James, 2000a; Berg et al.

2005).[2] This in itself suggests the need to collect more BDM data, as this current paper does, in order to determine if the finding of r_i greater than one with the BDM is as consistently replicable as the finding of r_i less than one when using the first price auction.

In light of the results discussed in this section of the paper, which already establish cross-institutional inconsistencies in risk preference parameter estimates, this current paper focuses on a particular institution (BDM) in an attempt to observe influences on choice data that are clearly *not* attributable to subjects' own risk preferences. Subjects' learning, as reflected in adoption of the institution's dominant strategy, and subjects' responses to variations in exogenously imposed payoff transformation regimes are both possible indicators that something other than risk preferences might be influencing subjects' behavior in BDM.

3 Research questions

The experiments in this paper are intended to address the following questions.

Question 1a: What estimates are obtained for r_i using data from the selling version of BDM?
Question 1b: What estimates are obtained for r_i using data from the buying version of BDM?
Question 2: Do the respective estimates from each version of BDM *change* over time?
Question 3: What level of revelation (equivalently, use of the dominant strategy) is observed in the BDM?
Question 4: Does that level of revelation *change* over time?
Question 5: Is misrevelation associated with deviation from risk neutrality?
Question 6: Is misrevelation associated with parameter instability?
Question 7: What effect do exogenously imposed non-linear payoff transformations have on choice in the BDM?

4 Experimental design

In addition to expanding the body of work on BDM in terms of quantity, this paper also takes a qualitatively different approach through several key design features. These design features may allow for the gathering of data which might suggest explanations for asymmetry in estimates of r_i that do not necessarily rely on risk attitudes.

First, in this design the time series of individual choices is broken into four distinct regimes: an initial baseline, two treatments, and a final baseline. This provides for distinct early and late baseline estimates of r_i made according to pre-determined break points. This allows an assessment of the stability of each individual's r_i estimate over time, which in turn may provide clues as to the nature of asymmetry in such estimates.

[2] Interestingly, Isaac and James (2000a) and Berg et al. (2005) in replicating the results of Cox et al. (1988) using the first price auction do so using the same subjects who generated risk-seeking data within the BDM.

Second, the experimental design provides a way to measure the subjects' understanding of BDM, and their possible learning over time. This is accomplished by means of including periods where the lottery is degenerate (the probability of a given payoff is one). When the lottery is degenerate BDM becomes a second price auction for an induced value commodity object, and the experimenter is able to observe whether the subject acts as if she both understands the institution and truthfully reveals her induced value, or not. The data from these periods will be directly comparable to earlier work in this area, including Attiyeh et al. (2000), Cox et al. (1996), and Isaac and James (2000b), all of which document subjects' convergence over time towards true revelation of object value in incentive compatible institutions.

Third, the treatment regimes that comprise the middle part of the design implement payoff transformations.[3] The most closely related work in this regard is that of Cox, Smith, and Walker, who as part of their larger research program ran some experiments wherein intra-experiment payoffs were translated into U.S. dollar payoffs by means of one or the other of the following functions:

$$\text{U.S. dollars} = (\text{Experimental dollars})^{1/2}$$
$$\text{U.S. dollars} = (\text{Experimental dollars})^{2}$$

This current paper implements each of these payoff transformations during respective treatment conditions for every subject in the pool. Doing so allows the gathering of choice data generated under distinct payoff environments, but which—given a maintained hypothesis of CRRA expected utility maximization—should nonetheless be consistent with a stable estimate of r_i for each subject. The effects of such transforms in a BDM context can also be directly compared to the results obtained by Cox, Smith, and Walker in the first price auction.[4] Fourth, in addition to the new design features just outlined, this paper also implements BDM in both selling and buying versions, as done by Kachelmeier and Meheta (1992).

Taken together, these design features might provide for some clues as to the nature of asymmetry in risk preference parameter estimates obtained from BDM data. For instance, were the estimates from the early and late baseline to differ, it would raise questions about the interpretation of an r_i estimate as a behavioral constant. And if time variation in subjects' use of their dominant strategy of truthful revelation were observed, this might call into question the extent to which choices from all periods reflected the same behavior. Also, differences in the pattern of estimates across the buying and selling versions of BDM might prove informative. Furthermore, one might be able to combine such observations to suggest still other conclusions.

[3] It should be noted that the procedures discussed in this regard are distinct from the lottery payoff procedure developed by Roth and Malouf (1979) and tested by Berg et al. (1986).

[4] Other previous work within this part of the literature might allow interesting comparisons with the work in this paper. For example, James and Isaac (2000) examine the effect of convex (specifically "beat the market") incentive contracts for traders in an experimental asset market. While the transformations used here match those used by Cox, Smith, and Walker rather than those used by James and Isaac, there might still be some value in comparing the effect of non-linear payoff transformations in individual choice versus market settings.

The experiments engaged in here assume as a maintained hypothesis—for argument's sake and for comparability with other work—that the subjects have CRRA utility functions defined over U.S. dollar income (as opposed to wealth) expressed as:

$$u(y) = y^{r_i}$$

If in addition to assuming $u(y) = y^{r_i}$ as a maintained hypothesis, one implements payoff transformations of the form:

$$\text{U.S. dollars} = (\text{Experimental currency})^{\alpha}$$

during the experiment, then utility defined over experimental currency is $u(c) = c^{\alpha r_i}$. A subject's valuation of a two-state lottery with a zero payoff in the low state can then be written as:

$$u(\text{CE}) = \text{CE}^{\alpha r_i} = \mathrm{p_{high}}(\mathrm{c_{high}})^{\alpha \mathrm{r_i}} + (1 - p_{\text{high}})(0)^{\alpha \mathrm{r_i}}$$

Taking natural logs and dividing both sides by αr_i yields:

$$\ln(\text{CE}) = \ln(p_{\text{high}})/\alpha r_i + \ln(c_{\text{high}})$$

One can then append multiplicative dummies (to account for each of the payoff transformation treatments and the second baseline) and a standard econometric error term to derive the estimating equation:

$$\ln \text{CE} = a + b_1 \ln(p_{\text{high}}) + b_2(D_{\text{convex}}) + b_3(D_{\text{concave}}) + b_4(D_{\text{second}}) + u_{it} \qquad (1)$$

where:

CE = the certainty equivalent for a given lottery, as written down by the subject
a = constant term, which is restricted to equal the natural log of the high payoff in the lottery, i.e. $\ln(c_{\text{high}})$
$\ln(p_{\text{high}})$ = natural log of the probability of the high payoff in the lottery
D_{convex} = a multiplicative dummy taking on the value $\ln(p_{\text{high}})$ when the convex payoff transformation is in effect (and 0 otherwise)
D_{concave} = a multiplicative dummy taking on the value $\ln(p_{\text{high}})$ when the concave payoff transformation is in effect (and 0 otherwise)
D_{second} = a multiplicative dummy taking on the value $\ln(p_{\text{high}})$ during the second baseline (and 0 otherwise)
u_{it} = error term satisfying the OLS assumptions.

Estimating (1) in time series, then solving for r_i from $b_1 = 1/r_i$ gives an estimate of the subject's r_i during the first baseline; similarly, $r_i = 1/(b_1 + b_4)$ gives an estimate of the subject's r_i during the second baseline. Also it should be noted that $b_2 = b_3 = (1 - \alpha)/\alpha r_i$. Since $\alpha = 2$ during the convex payoff transformation treatment, b_2 should be negative; and given $\alpha = 1/2$ during the concave payoff transformation treatment, b_3 should be positive.

The experiments were run so as to collect data suitable for estimating (1). This involved implementing BDM in both buying and selling forms. The operation of each version of BDM will now be outlined.

Selling version of BDM

In each round, the subject starts out owning the proceeds to a lottery with p_{high} of a two dollar payoff, and $(1 - p_{\text{high}})$ of a zero payoff. The subject then writes down the value she places on the rights to those lottery proceeds. The experimenter then draws from a uniform distribution to determine whether the subject will keep the rights to the proceeds of the lottery. If the number drawn is less than the subject's offer, the subject keeps the rights to the proceeds of the lottery. If the number drawn is greater than the subject's offer, the subject is held to have sold the rights to the lottery proceeds to the experimenter for an amount equal to the draw from the uniform distribution. Next, and finally, the lottery took place. p_{high} was varied round by round.

Buying version of BDM

In each round, the subject has an opportunity to purchase the rights to the proceeds of a lottery with p_{high} of a two dollar payoff, and $(1 - p_{\text{high}})$ of a zero payoff. The subject first writes down the value she places on the rights to those lottery proceeds. The experimenter then draws from a uniform distribution. If the number drawn is less than the subject's offer, the subject is awarded the rights to the proceeds of the lottery and pays to the experimenter an amount equal to the draw from the uniform distribution. If the number drawn is greater than the subject's offer, the subject does not purchase the rights to the lottery proceeds from the experimenter. Next, and finally, the lottery takes place. p_{high} was varied round by round. [5]

Each individual choice experiment used either the selling version of BDM or the buying version of BDM, but not both. All experiments using either version of BDM had 52 rounds. For all subjects, the first thirteen rounds were run with experimental currency earnings translated into U.S. dollars one for one. For half of the subject pool, the second thirteen periods used the payoff transformation, U.S. dollars = (Experimental currency)$^{1/2}$, while the third thirteen rounds used the payoff transformation, U.S. = (Experimental currency)^2. The other half of the subject pool had the order of the payoff transformation treatments reversed. Finally, all subjects were returned to the original one for one payoff conversion for the final thirteen periods.

Also, it should be noted that in each thirteen round segment (first baseline, first treatment, second treatment, second baseline) there was one round where p_{high} equaled either zero or one. In these rounds, the value of the lottery was independent of risk

[5] The buying version of BDM requires that a subject have working capital with which to bid. For 14 of the 28 subjects, their initial working capital was set at \$20.00 (which was theirs to keep). The other 14 subjects started with working capital of \$3.00 (again, which was theirs to keep). It is also possible for a subject to then run out of working capital. Subjects were told in advance that they would cease bidding immediately and for the remainder of the experiment if their working capital balance went negative. This situation never arose in the course of the experiments, however. The results reported for the buying version of BDM in this paper do not classify estimates by initial endowment, but such a classification is available upon request.

preference, and was equal to the payoff received with certainty. This enables the experimenter to determine if a subject is behaving as if she both understands the BDM institution and is revealing her true valuation of the lottery, or not.

The experiments each took around ninety minutes to run. Subjects were recruited from students in upper division economics classes at the University of Arizona and at Fordham University. Upon showing up, subjects were paid a $7.00 show-up fee and began to read the instructions (included as an Appendix).

Prior to commencing the experiment the subjects read the instructions and completed two practice rounds (without monetary payment) during which the operation of BDM was reviewed and questions were fielded. The instructions were written with the objective of giving BDM its best shot at success. That is to say, the instructions expressly inform the subjects that it is in their interest to truthfully report their value, and this point is then illustrated in the instructions with cases showing the harmful effects of either over-bidding or under-bidding. Upon completion of the experiment, subjects were paid their earnings from the experiment. These ranged from $20.00 to $88.00.

5 Results

5.1 Estimation results

The individual subject estimations are reported below, first for the selling version of BDM, then for the buying version of BDM. For all subjects, the two rounds where $p_{\text{high}} = 0$ were dropped from estimation, due to the behavior of the natural log evaluated at zero. This leaves a time series with $T = 50$. Beyond that, in the selling version, 7 of the 28 subjects offered 0 in some of the other rounds; in the buying version, 16 of the 28 subjects bid 0 in some of the other rounds. In Tables 1 and 2, the estimates obtained when also dropping those rounds for those subjects are reported. Alternative measures based on censored data techniques are also possible; estimates obtained by replacing 0 with 0.01 as a robustness check are noted for the selling version in endnote 6, and for the buying version in endnote 8.

Also, it should be noted that the calculation of R^2 is different when a constant term restriction is in place, as it is here as required by the theoretical derivation of the estimating equation. In this case, R^2 can be less than zero for a particularly poorly fitting regression.

Tables 1 and 2 present individual estimations that can be interpreted on a case-by-case basis, and doing so reveals a wide range of behavior. The estimates obtained for individuals can also be examined as a cross-section, and doing so provides interesting new data addressing research questions noted earlier in this paper.

First, the stability of estimates of the parameter r-across studies and across time within this study—using BDM data is addressed. Earlier findings of average estimates greater than one using the selling version of BDM are replicated here, but with a key new piece of information. The experimental design allows distinct early and late baseline sessions (separated by treatment sessions) and hence allows distinct early and late estimates of r. The mean (median) of the early estimates is 1.34 (1.20), but the

Table 1 Individual estimates from selling version of BDM

Subject	Coeff. on p_{high} (Std. error)	Coeff. on D_{convex} (Std. error)	Coeff. on D_{concave} (Std. error)	Coeff. on D_{second} (Std. error)	T	R^2
1	.6602	−.2769	.0302	.0388	50	.80
	(.0640)	(.0909)	(.0909)	(.0909)		
2	2.5303	−.4432	−1.2478	.1240	50	.57
	(.3207)	(.4535)	(.4535)	(.4535)		
3	1.0068	−.2808	.7739	−.1343	42	.92
	(.0733)	(.1036)	(.1036)	(.1036)		
4	1.1447	−.2380	−.2226	.1360	50	.91
	(.0670)	(.0949)	(.0949)	(.0949)		
5	.7045	−.0187	.5589	.4116	50	.94
	(.0456)	(.0645)	(.0645)	(.0645)		
6	1.7141	−.3416	−.5576	.2170	49	.83
	(.1661)	(.2349)	(.2628)	(.2349)		
7	.3639	.0129	.0283	.0908	50	.59
	(.0528)	(.0747)	(.0747)	(.0747)		
8	.4311	.4144	.3766	.6346	50	.87
	(.0708)	(.1002)	(.1002)	(.1002)		
9	1.3343	.3577	.0557	.3631	50	.88
	(.1248)	(.1765)	(.1765)	(.1765)		
10	.7625	.4147	.3390	.2854	50	.81
	(.1103)	(.1561)	(.1561)	(.1561)		
11	.6012	−.2615	.5256	1.1889	48	.34
	(.2532)	(.3268)	(.3268)	(.3581)		
12	1	0	0	0	50	1
	(2.47 e17)	(0)	(0)	(0)		
13	.3917	.6275	.4613	.5645	45	.55
	(.2303)	(.3192)	(.3192)	(.3192)		
14	.7103	.8395	.2732	1.0216	45	.66
	(.1387)	(.2784)	(.2195)	(.2784)		
15	.5211	.2845	.5764	.5606	50	.85
	(.0879)	(.1243)	(.1243)	(.1243)		
16	1.6180	−.3324	−.3326	.0335	40	.85
	(.0879)	(.1618)	(.1766)	(.1766)		
17	1.3576	.4678	1.0215	1.1982	50	.63
	(.3301)	(.4668)	(.4668)	(.4668)		
18	.9048	.1291	.1473	.1299	50	.81
	(.0933)	(.1320)	(.1320)	(.1320)		
19	.9542	−.1202	−.3597	−.0767	33	.64
	(.1673)	(.2367)	(.2241)	(.2678)		
20	1.2465	.1183	.1994	.5436	50	.73
	(.1645)	(.2326)	(.2326)	(.2326)		
21	.6840	1.0655	.3597	.9744	50	.64
	(.2462)	(.3483)	(.3483)	(.3483)		
22	.4704	0	0	−.0096	50	−17.31
	(.1135)	(.1606)	(.1606)	(.1606)		
23	.6737	.2022	.1287	0.6096	50	.06
	(.3581)	(.3581)	(.3581)	(.3581)		

(*Continued on next page*)

Table 1 (*Continued*)

Subject	Coeff. on p_{high} (Std. error)	Coeff. on D_{convex} (Std. error)	Coeff. on D_{concave} (Std. error)	Coeff. on D_{second} (Std. error)	T	R^2
24	.9884	.3097	.0722	.2542	50	.69
	(.1511)	(.2137)	(.2137)	(.2137)		
25	.9887	−.1229	−.0832	−.0244	50	.93
	(.0482)	(.0682)	(.0682)	(.0682)		
26	1.4501	.1321	−.3543	.2031	50	.80
	(.1720)	(.2432)	(.2432)	(.2432)		
27	.4112	.9903	.7115	.4023	50	.34
	(.4158)	(.5880)	(.5880)	(.5880)		
28	.6053	.2917	.3199	.3394	50	.77
	(.0902)	(.1276)	(.1276)	(.1276)		

Estimating equation: $\ln \text{CE} = a + b_1 \ln(p_{\text{high}}) + b_2(D_{\text{convex}}) + b_3(D_{\text{concave}}) + b_4(D_{\text{second}}) + u_{it}$.

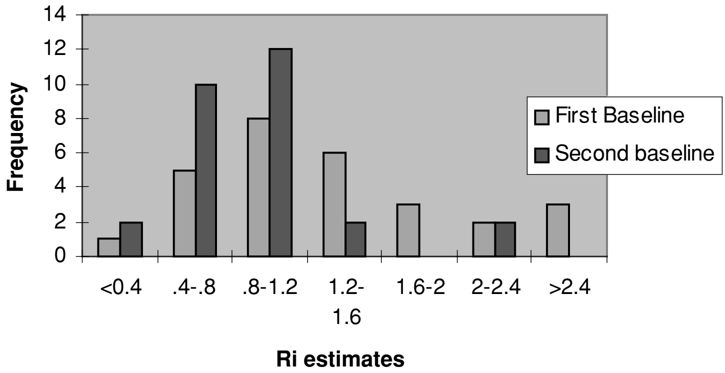

Fig. 1 Plot of individual estimates from the selling version of BDM

mean (median) of the late estimates is .92 (.91).[6] Graphically, the empirical densities are shown in Fig. 1.

The plot of the estimates of the individual subjects' r_i parameters appears to decrease in mean and decrease in variance over time.

[6] Instead of dropping from estimation those observations where the subject reported a certainty equivalent equal to 0, as was done for the estimates in the table and in the frequency plot, one can replace certainty equivalents of 0 with 0.01 as a robustness check. Doing so one obtains a mean (median) of 1.21 (1.01) in the early baseline, and a mean (median) of 0.85 (0.78) in the late baseline. This approach, though crude, is useful. It gives a lower bound in estimating r_i; since it fills in the missing data with the lowest certainty equivalent that is both possible in the experiment and does not map to an undefined value of the natural log function. More nuanced censored data techniques would yield higher estimates. (For instance, if one interprets reported certainty equivalents of zero for lotteries with positive, but low, expected value as showing the subject declining the effort of deciding whether the lottery is really worth 0.03 or 0.05 to her, rather than showing that the subject actually values the lottery at zero, then in principle some missing observations should be filled in with values greater than 0.01.)

Table 2 Individual estimates from buying version of BDM

Subject	Coeff. on p_{high} (Std. error)	Coeff. on D_{convex} (Std. error)	Coeff. on $D_{concave}$ (Std. error)	Coeff. on D_{second} (Std. error)	T	R^2
1	1.1532 (0.1087)	0.2110 (0.1537)	−0.0616 (0.1537)	0.2436 (0.1537)	50	.85
2	2.0186 (0.1094)	−1.5391 (0.2397)	−1.8767 (0.3236)	−2.0186 (0.3563)	36	.90
3	0.8497 (0.1681)	0.2365 (0.2377)	−0.0791 (0.2377)	−0.2922 (0.2377)	50	.44
4	1.0230 (0.0410)	−0.0235 (0.0580)	0.0917 (0.0580)	0.0356 (0.0580)	50	.96
5	1 (0.0711)	0.0595 (0.1005)	0.3585 (0.1005)	−0.0085 (0.1005)	50	.91
6	1.1887 (0.0829)	−0.4025 (0.1172)	−0.0807 (0.1172)	−0.4219 (0.1172)	50	.85
7	1.1772 (0.0897)	−0.1156 (0.1419)	0.0283 (0.1419)	−0.3007 (0.1800)	46	.86
8	1.1663 (0.1476)	0.1608 (0.1904)	0.0630 (0.2087)	0.2500 (0.1904)	48	.82
9	−0.3249 (0.2699)	0.5439 (0.3817)	0.3403 (0.3817)	0.3613 (0.3817)	50	.04
10	0.9263 (0.1286)	0.8672 (0.1819)	0.1770 (0.1819)	0.2798 (0.2929)	47	.73
11	1 (0.0027)	−0.0015 (0.0038)	0 (0.0047)	0 (0.0038)	46	.99
12	0.9357 (0.0705)	0.0906 (0.0997)	0.1557 (0.0997)	0.0032 (0.0997)	50	.90
13	1.2841 (0.1026)	−0.3434 (0.1451)	0.2923 (0.1451)	−0.3503 (0.1451)	50	.87
14	1.4160 (0.0668)	−0.3639 (0.0944)	0.2398 (0.1159)	−0.2799 (0.1159)	44	.92
15	0.2420 (0.0688)	−0.2706 (0.1829)	−0.2420 (0.1441)	−0.2420 (0.1829)	33	.29
16	1.7092 (0.2380)	−0.8373 (0.3598)	−1.1054 (0.3461)	−0.3300 (0.3367)	35	.58
17	0.8538 (0.0430)	0.2904 (0.0681)	0.1791 (0.0864)	0.2380 (0.0681)	46	.93
18	1 (0)	0 (0)	0 (0)	0 (0)	50	1.0
19	1.2910 (0.1033)	0.0466 (0.1333)	0.3748 (0.1333)	0.1810 (0.1333)	49	.92
20	0.9272 (0.0374)	0.0056 (0.0529)	−0.0863 (0.0529)	0.0566 (0.0529)	50	.96
21	0.9430 (0.0537)	−0.0686 (0.0850)	0.4438 (0.0850)	0.0494 (0.0850)	47	.93
22	1.1580 (0.0663)	0.0941 (0.1049)	−0.0883 (0.0937)	−0.1375 (0.0937)	49	.92
23	1.2666 (0.0732)	−0.3482 (0.1035)	0.1629 (0.1035)	0.1337 (0.1035)	50	.91

(*Continued on the page*)

Table 2 (*Continued)*

Subject	Coeff. on p_{high} (Std. error)	Coeff. on D_{convex} (Std. error)	Coeff. on $D_{concave}$ (Std. error)	Coeff. on D_{second} (Std. error)	T	R^2
24	1.4360 (0.1062)	−1.4360 (0.2066)	−0.4434 (0.1398)	−1.4360 (0.2494)	41	.86
25	1.3028 (0.1990)	−0.8917 (0.2368)	−0.2631 (0.3221)	−0.1153 (0.2215)	41	.76
26	1.1084 (0.0355)	−0.0010 (0.0502)	0.1653 (0.0502)	0 (0.0502)	50	.98
27	1.0545 (0.1501)	0.0048 (0.2048)	−0.5361 (0.3318)	0.0308 (0.1883)	43	.74
28	1.0245 (0.1097)	0.9543 (0.3061)	0.4919 (0.1818)	1.0709 (0.3061)	40	.56

Estimating equation: $\ln \text{CE} = a + b_1 \ln(p_{\text{high}}) + b_2(D_{\text{convex}}) + b_3(D_{\text{concave}}) + b_4(D_{\text{second}}) + u_{it}$.

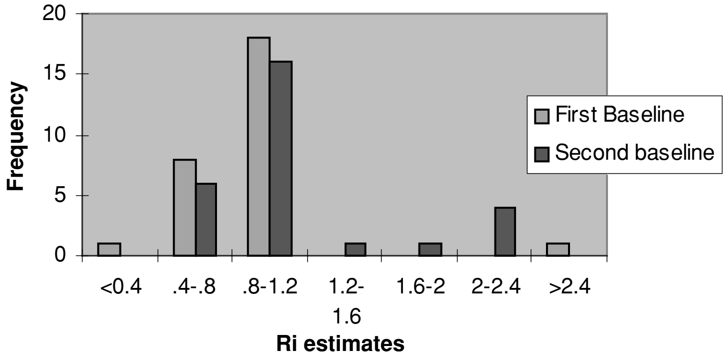

Fig. 2 Plot of individual estimates from the buying version of BDM

Using the buying version of BDM, the mean (median) of the early estimates is 0.88 (0.88), but the mean (median) of the late estimates is 1.09 (0.99).[7,8] The individual estimates are plotted in Fig. 2.

[7] In addition to the issue of certainty equivalents being reported as zero, already raised in connection with the data from the selling version of BDM, these data from the buying version of BDM introduce a new question. How does one handle regression estimates that imply an r_i of infinity (that is, infinitely risk-seeking behavior)? This is observed for 4 subjects in the late baseline of the buying version of BDM. For these subjects, an estimate of "2" is substituted for infinity when calculating statistics such as the population mean, and for consistency is also used in the frequency plot of estimates. This captures the qualitative nature of the results, without skewing the mean to the point of meaninglessness. (Also, the median is reported as an alternative measure that does not suffer from this sensitivity.)

[8] As before, one can perform a robustness check by re-estimating the equation with 0.01 replacing 0 for periods when the subject reported a certainty equivalent of 0 (and thus not be forced to drop from the estimation those periods with a dependent variable equal to the natural log of 0). Doing so one finds that the mean (median) of the early estimates is 0.70 (0.85), and that the mean (median) of the late estimates is 0.84 (0.78). Again, these represent the lowest possible values for these estimates; replacing certainty equivalents of 0 with any permissible value other than 0.01 yields higher estimates.

Table 3 Tally of hypothesis test results on subjects' b_4 coefficients

	Selling version	Buying version
Number of subjects showing statistically significant shift to higher r_i	0	6
Number of subjects showing statistically significant shift to lower r_i	13	2
Number of subjects showing no statistically significant shift either way	15	20

Note: Significance is calculated for 95% level.

Hypothesis tests of whether individual subjects' second baseline coefficients (from Tables 1 and 2) are respectively significantly different from zero can shed some light on whether or not parameter estimates shift from the early baseline to the late baseline. Tallying the results of those hypothesis tests generates Table 3.

We see from this table that parameter estimates exhibit a statistically significant shift away from the risk seeking end of the scale for 13 of 28 subjects in the selling version of BDM, while the other 15 subjects in that institution do not register a significant change one way or the other. For the buying version we then note that of the 28 subjects in that institution, 6 show a statistically significant shift away from the risk averse end of the scale, 2 show a statistically significant shift in the opposite direction, and the other 20 do not show a statistically significant shift in either direction.[9]

Furthermore, the distributions of estimates from the different versions of BDM appear to be converging over time. Under the *early* baseline, a Wilcoxon rank-sum test rejects that the distribution of estimates from the selling version and the distribution of estimates from the buying version are the same (calculated $z = 3.408$, critical $z = 1.6$ for $\alpha = .05$). But under the *late* baseline, a Wilcoxon rank-sum test fails to reject that the two distributions are the same ($z = 0.77$, critical $z = 1.6$ for $\alpha = .05$).

5.2 Confusion and learning

Since the data include four periods where the payoff of the lottery is known with certainty, an unambiguous appraisal of the revelation properties of the BDM—and a comparison with other dominant strategy institutions—is possible. Following Isaac and James (2000b) both exact revelation and approximate revelation are reported. Approximate revelation is here defined as within one cent of exact revelation; the approximate revelation series therefore contains the exact series as a subset. The

[9] It should also be noted here that changes in accumulated earnings during the experiment are unlikely to be a satisfactory explanation for the changes in risk preference parameter estimates observed in this experiment. The reason for this is that were one to nominate changes in wealth as causing risk preference changes in subjects with utility functions exhibiting increasing relative risk aversion (IRRA) as an explanation of the results in the selling version, one would then have to explain why a similar pattern of behavior is *not* observed in the buying version. If anything, subjects in the buying version would be more accurately characterized as exhibiting constant relative risk aversion (CRRA) or even decreasing relative risk aversion (DRRA). A changes-in-wealth based explanation would have to incorporate some auxiliary theory explaining the disparity observed across institutions. (For that matter some economists might object to invoking IRRA at all, on theoretical grounds.)

Table 4 Percentage of subjects using dominant strategy

	Selling version exact revelation	Selling version approximate revelation (within one cent of exact)	Buying version exact revelation	Buying version approximate revelation (within one cent of exact)
1st Baseline	35.72%	50%	89.3%	92.9%
1st Treatment	42.86%	53.57%	89.3%	96.4%
2nd Treatment	46.43%	82.14%	71.4%	89.3%
2nd Baseline	53.57%	85.71%	71.4%	89.3%

respective percentages of the subject pool following the dominant strategy of revelation in each version of BDM are presented in Table 4.

Higher levels of exact revelation are observed here than in comparable rounds of earlier experiments from other dominant strategy institutions. For example, Cox et al. observe 35% exact revelation by period 20 of their second-price auction data, while Isaac and James (2000b) observe 19.44% exact revelation by period 20 when using the incentive compatible combinatorial auction; these can be compared to the 42.86% exact revelation observed in round 19 when using the selling version or 89.3% in round 19 when using the buying version.

It also happens that violations of the dominant strategy have a statistically significant cross-sectional relationship with deviations from risk neutrality. Specifically, one can represent a subject's adoption (or not) of the dominant strategy of truthful revelation with a dummy variable, and a subject's deviation from risk neutrality by the absolute value of the difference between that subject's estimated r_i and the risk neutral value, $r = 1$. Pooling the data from the selling and buying versions of BDM (for a total sample of $N = 56$), the cross-sectional relationship between these two series can then be modeled in a number of possible ways: ordinary least squares, probit analysis, or non-parametric statistics. All of these various approaches in this case yield the same results. First, increased absolute deviation from risk neutrality and failure to adopt the dominant strategy are positively correlated in cross-section, in the early baseline (for the Wilcoxon rank-sum test, calculated $z = 3.07$, critical $z = 1.6$). Second, when variation in the cross-section of estimates has diminished by the late rounds of the experiment (as illustrated and discussed in Section 5.1), this correlation disappears in the late baseline (for the Wilcoxon rank-sum test, calculated $z = 0.637$, critical $z = 1.6$).

Furthermore, one could also posit that those subjects who utilize the dominant strategy in the first baseline (who arguably understand the institution best) should exhibit the least change in their parameter estimates from the early baseline to the last baseline.[10] The argument here is that since those subjects are less confused, they have less to learn, and if learning is (at least partly) behind changes in parameter estimates, then all else being equal these subjects' parameter estimates should change less. This conjecture can be addressed by performing a Wilcoxon rank-sum test to determine whether the absolute change between the early and late baseline parameter estimates

[10] I would like to thank Svetlana Pevnitskaya for suggesting this approach to the data.

itself tends to differ between the group who utilized the dominant strategy in the first round, and those who did not. One finds that that the groups are in fact different: the parameter estimates are more stable for the group who used the dominant strategy in the first baseline (for the Wilcoxon rank-sum test, calculated $z = 2.35$, critical $z = 1.6$).

These findings are not inconsistent with the conjecture that at least some deviations from risk-neutrality in laboratory data—particularly in the early rounds, when those deviations are most pronounced—might be the result of confusion on the part of the subjects, rather than an indication of their risk preferences. These findings complement those of Isaac and James (2000a), who found that in own-subject controlled, cross-institutional comparisons that subjects who appeared the most risk averse in the first price sealed bid auction were likely to appear the most risk seeking in the selling version of BDM, while the subjects who were risk neutral in one institution were likely to be risk neutral in the other. One possible explanation for this result of Isaac and James is that the subjects who demonstrate risk preference parameter inconsistency across institutions—which happened also to be those far from risk neutrality—are in some way confused. The design in this present paper allows for a reappraisal of that conjecture based on a simple proxy for subject comprehension (adoption of the dominant strategy). The results seem to offer further support for that conjecture.

Interestingly, just as the respective distributions of risk preference parameter estimates from the two different institutions are statistically indistinguishable by the late rounds (as shown in Section 5.1), it is also the case that by the late rounds the occurrence of approximate revelation (bidding within one cent of valuation) across institutions is also essentially indistinguishable: 85.71% for the selling version versus 89.3% for the buying version. Furthermore, one might note that the buying version of BDM, which over the length of the experiment generates the higher and more stable adoption of the dominant strategy, is also the institution that generates the more stable risk preference parameter estimates. Conversely, the selling version generates a lower and less stable adoption of the dominant strategy, and also generates less stable parameter estimates.[11]

5.3 Payoff transformations

Finally, the data permit an assessment of the effect of payoff transformations on risky choice behavior and on our inferences about such behavior. If subjects maximize

[11] While the level of exact revelation in the buying version does drop from 89.3% in the first baseline to 71.4% in the last baseline, this does not necessarily invalidate the interpretation that high adoption of the dominant strategy is associated with stable risk preference parameter estimates. First of all, even when it drops to 71.4%, the level of exact revelation in the buying version is still higher than is ever achieved in the selling version (53.57% at its maximum). On that basis, and by comparison with other dominant strategy studies previously cited, exact revelation is still high. Second, the drop in exact revelation can be accounted for by a handful of subjects switching from exact revelation to approximate revelation—that is, changing their bids by *one cent*. This kind of fragility in exact revelation results is precisely the reason other papers investigating dominant strategy institutions have also reported approximate measures of revelation. Some examples of such past published papers include Cox et al. (1996), Attiyeh et al. (2000), and Isaac and James (2000b).

Table 5 Classification of treatment dummy coefficients

	Correct sign	Incorrect sign
	Selling version of BDM	
Statistically significant	12	12
Statistically insignificant	17	15
	Buying version of BDM	
Statistically significant	16	6
Statistically insignificant	13	21

the expectation of a given CRRA utility function defined over income, subject to the payoff transformation in effect at the time, then at a minimum one would predict that the coefficient on D_{convex} *should* be negative and that the coefficient on D_{concave} *should* be positive. One could go beyond that and make a *point prediction* for what each treatment coefficient should be given the payoff transformation and the subject's estimated r_i under the baseline. If the point prediction for a subject does not hold, expected utility maximization of an unchanged utility function subject to the treatment regime can be rejected. A fortiori, if the treatment coefficients are not significant, or do not even have the correct sign, one can tell by inspection that this intra-institutional (as distinct from cross-institutional) consistency check is being violated.

Organizing the treatment coefficient estimates by sign and significance, we have 56 observations (28 subjects participating in two treatments each) for each version of BDM. The observations for each version are divided into four cells in their respective parts of Table 5.

It appears that in general the payoff transformations do *not* have the effect predicted for subjects with CRRA utility. This finding, made in a BDM context, echoes the findings of Cox, Smith, and Walker when using the same payoff transformations in a first-price auction setting. The predominant lack of statistical significance contrasts with the results of James and Isaac, who found that non-linear payoff transformations clearly affected prices in an experimental asset market (though this latter comparison is confounded by a difference in the payoff transformation used).[12]

An alternative approach to analyzing the data generated under the non-linear payoff transformation treatments is to solve for the implied value of r_i for each subject in each treatment condition. To do this, one would solve the equation:

$$b_{\text{treatment}} = (1 - \alpha)/\alpha r_i$$

for r_i, taking α (set exogenously by the experimenter as discussed in the experimental design section) and $b_{\text{treatment}}$ (available in Tables 1 and 2) as given.

[12] The results in Table 6 on the effect of payoff transformation treatments show that while the transformations do not generally have the effect predicted for expected utility maximizing agents, such treatment coefficients as *are* statistically significant are more often of the correct sign for the buying version of BDM (16 out of 22) than was the case with case using the selling version of BDM (12 out of 24). This latter observation, viewed in conjunction with the facts that the data from the buying version show less violation of the dominant strategy of truthful revelation, makes one wonder whether the buying version might not be easier for the subjects to understand.

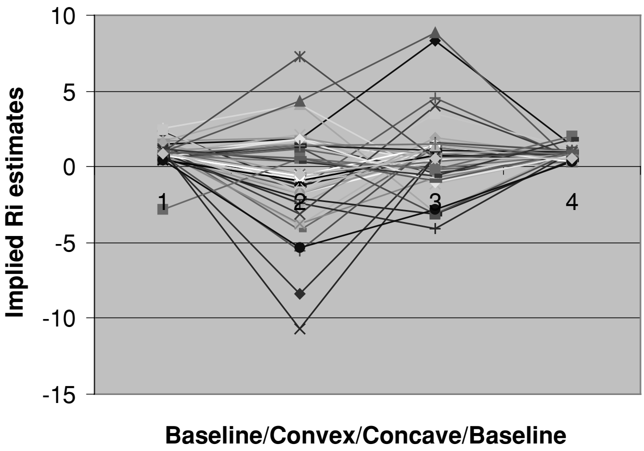

Fig. 3 Path of implied risk preference parameter estimates across treaments

The implied values of r_i shown in Fig. 3 are interesting for several reasons.[13] First, they are reminiscent of the anomalous results of Isaac and James, and of Berg, Dickhaut, and McCabe: if one believes the resultant implied values of r_i, then many subjects appear to switch from risk-seeking behavior to risk averse behavior (or vice versa) in moving from one treatment regime to the other. Interestingly, though, these results are obtained in an intra-institutional rather than cross-institutional setting. Second, for many subjects the values of r_i calculated in this fashion are *enormous*: in absolute value they are several times those estimated for the same subjects in either the early or late baseline regimes, and would seem to be implausible on this basis alone.

The question then becomes, what does one make of these implied values of r_i? One might suggest that the payoff transformations confuse the subjects and lead them to make errors they might not otherwise make in the absence of such transforms, and that in consequence the calculated values of r_i are not to be taken seriously.[14] That reasonable explanation raises another question, however: what kind of discounting of results do we employ when there is *not* a clearly labeled, exogenously imposed treatment involved? What happens when we move from one institution to another, rather than from one treatment to another within a single institution? What else might we be attributing to subject risk preferences that does not actually have anything to do with subject risk preferences? The payoff transformation treatments employed here could be interpreted as giving a demonstration of the kind of influence on implied risk preference parameters we might come across, but fail to see, in other settings.

[13] Outlying data associated with subjects whose responses yielded implied r_i parameters so extreme as to re-scale Fig. 3 into illegibility were deleted. Thus Fig. 3 tracks the implied r_i parameters across treatments for only 46 of the 56 total subjects. The implied r_i parameters for the other 10 subjects can be readily calculated using Tables 1 and 2.

[14] For instance, the less a subject responds to a payoff transformation, the closer the coefficient on the subject's dummy variable for that treatment will be to zero. As the coefficient approaches zero, the implied value of r_i approaches either positive or negative infinity (depending on whether a transformation exponent of less than or greater than one, respectively, is in place).

6 Conclusion

Four features of the data in this paper stand out both individually and in combination with each other. First, risk preference parameter estimates from the buying version and the selling version of BDM appear to converge towards each other over time. Second, the apparent convergence seems to be in the vicinity of risk neutrality (certainly it is nowhere near the parameter estimates from the first price auction). Third, deviation from risk neutrality appears to be correlated with violation of the dominant strategy of true revelation. Fourth, parameter instability also seems to be correlated with violation of the dominant strategy of true revelation. Taken together, these results raise the possibility that at least some deviations from risk neutrality in BDM might be the result of confusion on the part of the subjects, particularly in the early rounds, and not necessarily reflective of the risk preferences of subjects.

Acknowledgment The author gratefully acknowledges support from National Science Foundation grant SBR-9809741. I would like to thank John Dickhaut, Glenn Harrison, Charles Holt, Mark Isaac, Jamie Kruse, Bart Moore, Svetlana Pevnitskaya, Derrick Reagle, Tim Salmon, and session attendees at the Economic Science Association, 2003 North American Meeting and at the Southern Economic Association, 2004 Annual Meeting for their helpful comments. Additionally I would like to thank Tim Cason, Arthur Schram and two anonymous referees for providing helpful guidance throughout the review process.

References

Attiyeh, G., Franciosi, R., & Isaac, R. M. (2000). Experiments with the pivot process for providing public goods. *Public Choice*, *102*, 95–114.

Berg, J. E., Daley, L. A., Dickhaut, J. W., & O'Brien, J. R. (1986). Controlling preferences for lotteries on units of experimental exchange. *Quarterly Journal of Economics*, *101*(2), 281–306.

Berg, J. E., Dickhaut, J. W., & McCabe, K. (2005). Risk preference instability across institutions: a dilemma. *Proceedings of the National Academy of Sciences*, *102*(11), 4209–4214.

Cox, J. C., & Sadiraj, V. (2006). Small- and large-stakes risk aversion: implications of concavity calibration for decision theory. *Games and Economic Behavior*, *56*, 45–60.

Cox, J. C., Isaac, R. M., Cech, P., & Conn, D. (1996). Moral hazard and adverse selection in procurement contracting. *Games and Economic Behavior*, *17*, 147–176.

Cox, J. C., Roberson, B., & Smith, V. L. (1982). Theory and behavior of single object auctions. In L. Smith Vernon (Ed.), *Research in experimental economics*, vol. 2. Greenwich, CT: JAI Press.

Cox, J. C., Smith, V. L., & Walker, J. L. (1988). Theory and individual behavior of first-price auctions. *Journal of Risk and Uncertainty*, *1*, 61–99.

Dorsey, R. E., & Razzolini, L. (2003). Explaining overbidding in first price auctions using controlled lotteries. *Experimental Economics*, *6*, 123–140.

Friedman, M. (1953). The methodology of positive economics. In *Essays in positive economics*. Chicago: University of Chicago Press.

Harlow, V., & Brown, K. (1990). Understanding and assessing financial risk tolerance: a biological perspective. *Financial Analysts Journal*, *80*, 50–62.

Harrison, G. W. (1990). Risk attitudes in first-price auction experiments: a Bayesian analysis. *Review of Economics and Statistics*, *72*, 541–546.

Isaac, R. M., & James, D. (2000). Just who are you calling risk averse? *Journal of Risk and Uncertainty*, *20*(2), 177–187.

Isaac, R. M. & James, D. (2000b). Robustness of the incentive compatible combinatorial auction. *Experimental Economics*, *3*(1), 31–53.

James, D., & Isaac, R. M. (2000). Asset markets: how they are affected by tournament incentives for individuals. *American Economic Review*, *90*(4), 995–1004.

Kachelmeier, S. J., & Shehata, M. (1992). Examining risk preferences under high monetary incentives: experimental evidence from the People's Republic of China. *American Economic Review*, *82*(5), 1120–1141.

Rabin, M. (2000). Risk aversion and expected utility theory: A calibration theorem. *Econometrica*, *68*, 1281–1292.
Roth, A. E., & Malouf, M. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review*, *86*, 574–594.
Schorvitz, E. B (1998). Experimental tests of fundamental economic theories. Unpublished Ph.D. dissertation, University of Arizona.