Experimenting with Measurement Error: Comment *

by Daniel Friedman, Sameh Habib, Duncan James, and Brett Williams†

September 30, 2019

PRELIMINARY Version.

PLEASE DO NOT CIRCULATE without explicit permission

Gillen, Snowberg and Yariv (2019, henceforth GSY) present an estimation technique called ORIV intended to cope with measurement error, and argue that applying ORIV overturns conclusions obtained in some previous laboratory studies. We agree that not properly accounting for measurement error can invalidate inferences made from laboratory data, and that the ORIV technique can help cope with that problem. However, this Comment (a) shows that ORIV applied to GSY's data may actually reinforce previous conclusions regarding the elicitation of risk preferences, and (b) offers cautions for applying ORIV more generally.

GSY collect within-subject data in four risk preference elicitation tasks: Qualitative, Project, Lottery Menu, and Risk MPL. Table 6 of their paper displays cross-task correlations; for simplicity and comparability to other work, we focus on the central panel that involves elicited preferences expressed in terms of the coefficient of relative risk aversion (CRRA). That panel is reproduced in our Table 1 below, with Pearson correlations computed ignoring measurement error on the left ("Raw"), and those computed using ORIV on the right ("Corrected").

Normalization. For Risk MPL tasks, GSY use the formula $r = \frac{ln(0.5)}{ln(\frac{100}{V})} + 1$, where V is the normalized elicited value and r is the corresponding CRRA preference parameter. Raw elicited values for the 20-ball urn task lie in the interval [0, 100], but the given formula is undefined at

*We are grateful to Eric Snowberg and coauthors for giving us access to relevant data and to their STATA code.

†Friedman and Williams: University of California, Santa Cruz, Santa Cruz, CA 95064, dan@ucsc.edu and bwillia4@ucsc.edu; James: Economics Department, Fordham University dujames@fordham.edu; Habib: The Joint Committee on Taxation, sameh.habib@jct.gov

	Raw (censored)			Corrected (ORIV, censored)		
	Project	Qualitative	Lottery	Project	Qualitative	Lottery
Qualitative	0.20			0.36		
Lottery	0.27	0.24		0.55	0.38	
Risk MPL	0.18	0.07	0.22	0.37	0.10	0.42

Table 1: Estimated correlation of individual subjects' elicited coefficients of absolute risk aversion (CRRA) across tasks, using censored data as reported in GSY Table 6.

$V = 100$. GSY therefore normalize by setting $V \in [-10, 90]$ as the raw value minus 10; see lines 3, 9-10 of the STATA code reproduced in Appendix A.

Other Risk MPL observations come from a 30-ball urn with raw values in the interval [0,150]. GSY apparently intend a similar normalization here by subtracting 15 from each raw value and dividing by 1.5. However, apparently due to a typo, GSY's STATA code subtracts 5 instead of 15 before dividing by 1.5. A later line of code, perhaps intended to ameliorate the problem, replaces all values higher than 90 with 90; e.g., raw values of 140 and 150 both are coded as $V = 90$. See lines 4 and 12-14 of the STATA code reproduced in Appendix A. Table 2 below reports the correlation results when the apparently intended normalization is used for the 30-ball data.

	Raw (censored)			Corrected (ORIV, censored)		
	Project	Qualitative	Lottery	Project	Qualitative	Lottery
Qualitative	0.20			0.36		
Lottery	0.27	0.24		0.55	0.38	
Risk MPL	0.15	0.07	0.19	0.28	0.10	0.32

Table 2: Same as Table 1, but with the 30-ball urn normalized by subtracting 15 (instead of 5) and dividing by 1.5.

Censoring. For most correlation computations, GSY censor the Risk MPL data by converting the 442 observations revealing negative relative risk aversion (i.e., risk seeing) to zero relative risk aversion (risk neutrality), as as summarized in Table 3 below. Their reasoning is that the Project and Lottery tasks do not permit subjects to reveal negative relative risk aversion, and that censoring might help deal with that difference in task features.[1] Consequently, they use censored observations to compute Risk MPL vs Project and Risk MPL vs Lottery Menu correlations but not the Risk MPL vs Qualitative correlation.[2]

Table 4 reports updated correlations from Table 2 but using uncensored data. GSY (p. 22) say

[1]Many researchers use Spearman rank correlation rather than Pearson correlation to deal with differences across tasks in, e.g., the admissible range of revealed coefficients of relative risk preference. It seems to us an open question whether censoring improves the reliability of Pearson correlation coefficients.

[2]Had GSY not made this exception for Qualitative, which does permit subjects to reveal risk seeking preferences, they would have obtained Risk MPL - Qualitative correlations of 0.13 (raw) and 0.23 (ORIV).

	Uncensored			Censored		
	<0	0	>0	<0	0	>0
30 Urn	267	368	148	-	635	148
20 Urn	175	242	358	-	417	358

Table 3: Counts for risk-seeking (<), risk-neutral (0), and risk-averse (>0) elicited preferences in the 20-ball and 30-ball urn Risk MPL tasks. Left side ("Uncensored") shows actual counts, right side ("Censored") shows counts used in GSY.

that results are "qualitatively similar" with no censoring. To check, Table 5 reports the uncensored results when GSY's "subtract 5, divide by 1.5 and impose a ceiling at 90" normalization is used for the 30-ball urn data.

	Raw (uncensored)			Corrected (ORIV, uncensored)		
	Project	Qualitative	Lottery	Project	Qualitative	Lottery
Qualitative	0.20			0.36		
Lottery	0.27	0.24		0.55	0.38	
Risk MPL	0.08	0.07	0.11	0.14	0.10	0.15

Table 4: Same as Table 2, except that uncensored data are used in the Risk MPL tasks.

	Raw (uncensored)			Corrected (ORIV, uncensored)		
	Project	Qualitative	Lottery	Project	Qualitative	Lottery
Qualitative	0.20			0.36		
Lottery	0.27	0.24		0.55	0.38	
Risk MPL	0.09	0.07	0.11	0.14	0.10	0.14

Table 5: Same as Table 1, except that uncensored data are used in the Risk MPL tasks.

Sign convention. Running the GSY CRRA data through GSY's code, we obtained negative correlations of the Qualitative task with all other tasks. GSY's Qualitative task uses a scale from 0 to 10 with 0 being very risk-averse and 10 being very risk-seeking, the reverse of the convention for coefficients of risk aversion. Therefore, for all tables reported in this Comment, we reversed the Qualitative scale by simply subtracting all reported values from 10; this enabled us to reproduce GSY Table 6.

Range of Estimated Correlations. A key ORIV formula, given right before their Proposition 3, writes the corrected correlation between tasks X and Y as

$$\hat{\rho}^*_{XY} = \hat{\beta}^* \sqrt{\frac{\widehat{Cov}[X^a, X^b]}{\widehat{Cov}[Y^a, Y^b]}}. \tag{1}$$

In a moderate size samples, the sample covariance $\widehat{Cov}[Y^a, Y^b]$ of two observations of preference elicitation task Y may occasionally be much smaller (or even of different sign) than that of the other task, $\widehat{Cov}[X^a, X^b]$. Thus the formula may return absolute values greater than 1.0 (or even imaginary values). We found examples of this sort when we applied ORIV to simulated data.

The formal derivation of equation (1) and other ORIV formulas assumes that measurement errors embodied in X^a, X^b, Y^a and Y^b are uncorrelated. Although the assumption of uncorrelated measurement error is standard in theoretical developments, it is problematic in practice; see, for example, Bound et al. (1994), Rifkin (1995), and Day et al. (2004)). In their Section 4.4.2, GSY suggest experimental design practices that may reduce positively correlated measurement errors, and thus reduce residual bias in ORIV estimates. We note that negatively correlated measurement errors are also possible, e.g., if subjects consciously or unconsciously overcompensate for trembles in previous trials for a given elicitation task or think of the trials as portfolios. In that case, ORIV can overcorrect the attenuation bias even in large samples and return task pair correlation estimates whose absolute values are too high, and may even exceed 1.0.

Discussion. A spate of recent articles (e.g., Pedroni et al. (2017), Loomes and Pogrebna (2014), Friedman et al. (2019), Zhou and Hey (2018)) suggest that popular risk preference elicitation tasks produce inconsistent results, since cross-task correlations tend to be lower than one might expect to see if they indeed were measuring the same personal trait. A casual reader of GSY can easily get the impression that there is no such inconsistency and that, once measurement error is taken into account, the correlations are comfortably large.

The Tables presented above suggest that that impression is incorrect. With or without ORIV, the GSY data yield cross-task correlations that are similar to those obtained by previous researchers. Of course, positive correlations are higher with ORIV than without, but the cross-task correlations reported in Table 4, or even in Tables 2 or 5, lie in the range of 0.1-0.6, with the upper end of that range set by the correlation between the closely related tasks Project and Lottery. These correlation estimates line up fairly closely with those presented in previous articles.

We conclude that ORIV is a useful addition to the toolkit of experimental and other applied economists, but it must be used with caution.

Appendix: GSY Stata Code.

1. gen riskyUrn20Value = riskyUrn20MaxValue
2. gen riskyUrn30Value = riskyUrn30MaxValue
3. replace riskyUrn20Value = riskyUrn20Value - 10
4. replace riskyUrn30Value = (riskyUrn30Value - 5)/1.5

5. gen riskyUrn20RN = riskyUrn20Value
6. replace riskyUrn20RN = 50 if riskyUrn20Value > 50 & riskyUrn20Value = .
7. gen riskyUrn30RN = riskyUrn30Value
8. replace riskyUrn30RN = 50 if riskyUrn30Value > 50 & riskyUrn30Value = .

9. gen riskyUrn20CRRA = ln(0.5)/ln(100/riskyUrn20Value) + 1
10. replace riskyUrn20CRRA = ln(0.5)/ln(100/5)+1 if riskyUrn20Value < 5
11. gen riskyUrn30CRRA = ln(0.5)/ln(100/riskyUrn30Value) + 1
12. replace riskyUrn30CRRA = ln(0.5)/ln(100/5)+1 if riskyUrn30Value < 5
13. replace riskyUrn30CRRA = ln(0.5)/ln(100/90)+1 if riskyUrn30Value > 90

14. gen riskyUrn20CRRARN = riskyUrn20CRRA
15. replace riskyUrn20CRRARN = 0 if riskyUrn20CRRARN < 0
16. gen riskyUrn30CRRARN = riskyUrn30CRRA
17. replace riskyUrn30CRRARN = 0 if riskyUrn30CRRARN < 0