

ONLINE APPENDIX

LOTTERY-SPECIFIC EFFECTS AND HETEROGENEITY: A UNIFIED FRAMEWORK

JOSE APESTEGUIA[†], MIGUEL A. BALLESTER[‡], AND ÁNGELO GUTIÉRREZ-DAZA[§]

CONTENTS

Appendix A. Baseline Estimates by Treatment	2
Appendix B. Student- t Distribution	5
Appendix C. Additional Cross-validation Analysis	11

Date: October 2026.

*Financial support from FEDER/Ministerio de Ciencia e Innovación (Agencia Estatal de Investigación) through Grant PID2021-125538NB-I00 and through the Severo Ochoa Programme for Centers of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S), and from the EC through grant HORIZON-MSCA-2022-COFUND-101126600-SmartBRAIN3. The views in this paper do not necessarily reflect those of Banco de México.

[†]ICREA, Universitat Pompeu Fabra and Barcelona School of Economics. E-mail: jose.apestegui@upf.edu.

[‡]IKERBASQUE and Basque Country University. E-mail: miguelangel.ballester@ehu.eus.

[§]Banco de México. E-mail: angelo.gutierrez@banxico.org.mx.

APPENDIX A. BASELINE ESTIMATES BY TREATMENT

# Reference	Treatment	CRRA		CARA	
		Location	Scale	Location	Scale
1 Abdellaoui et al. (2011)	Ellsberg urns	-1.9508	8.7629	-0.0279	0.1236
2 Abdellaoui et al. (2011)	Natural events, hypothetical	-0.1111	1.4363	0.0001	0.0014
3 Abdellaoui et al. (2011)	Natural events, incentivized	-0.0761	2.6753	0.0002	0.0020
4 Beauchamp et al. (2020)	Hypothetical, list 0, EV hidden	-0.5882	1.7948	-0.0052	0.0133
5 Beauchamp et al. (2020)	Hypothetical, list 0, EV shown	-0.3207	1.0855	-0.0030	0.0091
6 Beauchamp et al. (2020)	Hypothetical, list -1, EV hidden	-0.2752	0.7382	-0.0030	0.0073
7 Beauchamp et al. (2020)	Hypothetical, list -1, EV shown	-0.2749	1.4571	-0.0024	0.0110
8 Beauchamp et al. (2020)	Hypothetical, list -2, EV hidden	-0.1134	0.8724	-0.0011	0.0080
9 Beauchamp et al. (2020)	Hypothetical, list -2, EV shown	-0.1612	1.2131	-0.0016	0.0097
10 Beauchamp et al. (2020)	Hypothetical, list +1, EV hidden	-0.5542	1.4414	-0.0048	0.0119
11 Beauchamp et al. (2020)	Hypothetical, list +1, EV shown	-0.5641	2.0532	-0.0044	0.0146
12 Beauchamp et al. (2020)	Hypothetical, list +2, EV hidden	-0.7400	1.7278	-0.0063	0.0127
13 Beauchamp et al. (2020)	Hypothetical, list +2, EV shown	-0.6809	1.7492	-0.0059	0.0141
14 Beauchamp et al. (2020)	Real, list 0, EV hidden	0.4200	0.7400	0.0040	0.0096
15 Beauchamp et al. (2020)	Real, list 0, EV shown	0.4221	0.6773	0.0040	0.0084
16 Beauchamp et al. (2020)	Real, list -1, EV hidden	0.3470	0.7339	0.0033	0.0089
17 Beauchamp et al. (2020)	Real, list -1, EV shown	0.2779	0.7545	0.0033	0.0108
18 Beauchamp et al. (2020)	Real, list -2, EV hidden	0.3605	0.7093	0.0034	0.0083
19 Beauchamp et al. (2020)	Real, list -2, EV shown	0.0751	0.6356	0.0013	0.0069
20 Beauchamp et al. (2020)	Real, list +1, EV hidden	0.5540	0.8288	0.0062	0.0144
21 Beauchamp et al. (2020)	Real, list +1, EV shown	0.4845	0.7434	0.0052	0.0107
22 Beauchamp et al. (2020)	Real, list +2, EV hidden	0.6595	0.7714	0.0066	0.0122
23 Beauchamp et al. (2020)	Real, list +2, EV shown	0.5016	0.7501	0.0053	0.0108
24 Bleichrodt et al. (2023)	Stadium capacity (ambiguous)	-0.0682	0.7546	0.0002	0.0026
25 Bleichrodt et al. (2023)	Canberra temperature (ambiguous)	0.0230	0.4494	0.0003	0.0017
26 Bleichrodt et al. (2023)	Paris temperature (ambiguous)	0.1598	0.7631	0.0016	0.0054
27 Bleichrodt et al. (2023)	Risk (known probabilities)	0.8870	4.2124	0.0024	0.0091
28 Bruhin et al. (2010)	Switzerland, 2003	0.1887	2.2426	0.0099	0.0925
29 Bruhin et al. (2010)	Switzerland, 2006	0.1509	2.0336	0.0072	0.0705
30 Bruhin et al. (2010)	China, 2005	-0.7136	8.5101	-0.0272	0.5369
31 Diecidue et al. (2015)	Study 1, hypothetical	0.2025	0.3227	0.0017	0.0023
32 Diecidue et al. (2015)	Study 1, real	1.1910	1.1428	0.0212	0.0221
33 Diecidue et al. (2015)	Study 2	-0.2661	1.3610	-0.0037	0.0361
34 Enke and Graeber (2023)	MTurk, baseline	-0.2121	1.3322	-0.0055	0.1027
35 Enke and Graeber (2023)	MTurk, ROCL	-0.0940	1.4572	0.0109	0.1279
36 Enke and Graeber (2023)	MTurk, high default	-0.2304	1.2447	-0.0076	0.0924

Continued on next page

Continued from previous page

# Reference	Treatment	CRRA		CARA	
		Location	Scale	Location	Scale
37 Enke and Graeber (2023)	MTurk, low default	0.0491	1.1528	0.0224	0.1266
38 Enke and Graeber (2023)	Prolific, hypothetical (BDM)	-0.5529	1.6997	-0.0357	0.1174
39 Fan et al. (2019)	1 stage	-1.6344	5.5549	-0.0814	0.2415
40 Fan et al. (2019)	2 stages, sequential	-3.6188	8.6552	-0.1851	0.3204
41 Fan et al. (2019)	2 stages, simultaneous	-3.7674	8.6028	-0.2018	0.3109
42 Fan et al. (2019)	3 stages, sequential	-3.0249	5.5598	-0.1629	0.2361
43 Fan et al. (2019)	3 stages, simultaneous	-3.3283	5.2957	-0.1916	0.2578
44 Fehr-Duda et al. (2010)	China, 2005	0.1303	2.2572	0.0029	0.0251
45 Gonzalez and Wu (2022)	1999 cohort	1.6955	4.5491	0.0087	0.0222
46 Gonzalez and Wu (2022)	2022 cohort	0.2396	3.7658	0.0011	0.0172
47 l'Haridon and Vieider (2019)	h_australia	0.3090	3.1376	0.0185	0.1022
48 l'Haridon and Vieider (2019)	h_belgium	-0.0241	2.7275	0.0037	0.0956
49 l'Haridon and Vieider (2019)	h_brazil	-0.2740	3.3920	-0.0047	0.1051
50 l'Haridon and Vieider (2019)	h_cambodia	-1.3207	5.5883	-0.0306	0.2028
51 l'Haridon and Vieider (2019)	h_chile	0.0822	3.6799	0.0124	0.1287
52 l'Haridon and Vieider (2019)	h_china	0.0564	2.2196	0.0050	0.0788
53 l'Haridon and Vieider (2019)	h_colombia	-0.4773	5.1923	-0.0054	0.1661
54 l'Haridon and Vieider (2019)	h_costa_rica	-0.4070	5.2572	-0.0010	0.1900
55 l'Haridon and Vieider (2019)	h_czech_republic	0.1273	1.2881	0.0076	0.0469
56 l'Haridon and Vieider (2019)	h_ethiopia	-1.1891	5.1890	-0.0333	0.1716
57 l'Haridon and Vieider (2019)	h_france	-0.1426	3.4847	0.0011	0.1176
58 l'Haridon and Vieider (2019)	h_germany	0.0628	3.2046	0.0094	0.1072
59 l'Haridon and Vieider (2019)	h_guatemala	-1.8762	9.1283	-0.0318	0.2957
60 l'Haridon and Vieider (2019)	h_india	-0.4926	3.9433	-0.0071	0.1487
61 l'Haridon and Vieider (2019)	h_japan	-0.0099	2.0336	0.0028	0.0688
62 l'Haridon and Vieider (2019)	h_kyrgyzstan	-0.7413	4.1787	-0.0171	0.1478
63 l'Haridon and Vieider (2019)	h_malaysia	-0.2044	4.1576	0.0017	0.1768
64 l'Haridon and Vieider (2019)	h_nicaragua	-4.6337	14.3129	-0.0897	0.5094
65 l'Haridon and Vieider (2019)	h_nigeria	-5.8129	17.8487	-0.0986	0.6345
66 l'Haridon and Vieider (2019)	h_peru	-1.7896	8.2612	-0.0385	0.2970
67 l'Haridon and Vieider (2019)	h_poland	-0.0333	2.3142	0.0021	0.0818
68 l'Haridon and Vieider (2019)	h_russia	-0.1897	2.8165	-0.0054	0.0953
69 l'Haridon and Vieider (2019)	h_saudi_arabia	-0.5830	3.1551	-0.0179	0.0951
70 l'Haridon and Vieider (2019)	h_south_africa	-0.3016	4.5899	-0.0015	0.1766
71 l'Haridon and Vieider (2019)	h_spain	0.0581	1.8499	0.0059	0.0653
72 l'Haridon and Vieider (2019)	h_thailand	-0.7040	4.7492	-0.0116	0.1513
73 l'Haridon and Vieider (2019)	h_tunisia	-0.6768	6.9041	-0.0061	0.2456
74 l'Haridon and Vieider (2019)	h_uk	-0.7662	2.9428	-0.0236	0.0937

Continued on next page

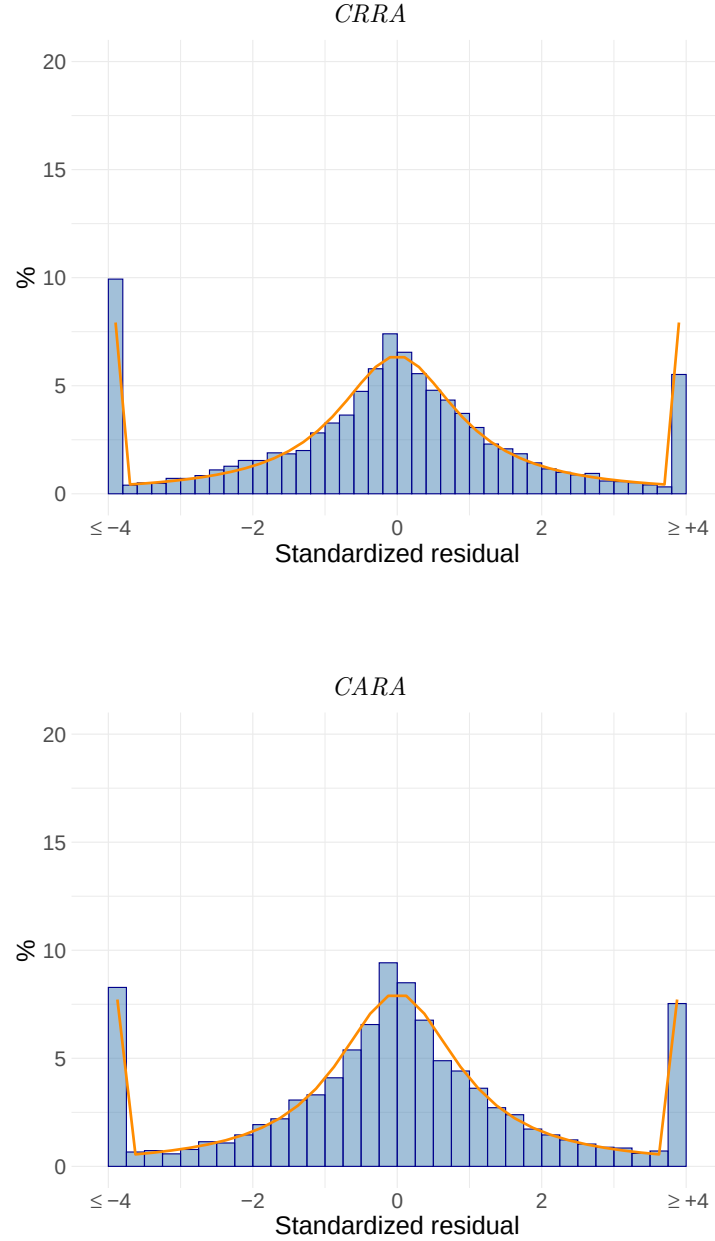
Continued from previous page

# Reference	Treatment	CRRA		CARA	
		Location	Scale	Location	Scale
75 l'Haridon and Vieider (2019)	h_usa	0.0653	2.6816	0.0073	0.0841
76 l'Haridon and Vieider (2019)	h_vietnam	-0.4731	3.3548	-0.0141	0.1100
77 Kontek and Birnbaum (2019)	Study 1 (Poland)	-0.0816	2.4906	0.0006	0.0136
78 Kontek and Birnbaum (2019)	experiment_2_pol	-0.1368	0.8213	-0.0002	0.0052
79 Kontek and Birnbaum (2019)	experiment_2_usa	0.0550	1.5647	0.0080	0.0334
80 Hajimoladarvish, N. (2018)	Compound lotteries	-1.1267	5.1185	-0.0290	0.1579
81 Hajimoladarvish, N. (2018)	Simple lotteries	2.1466	5.1453	0.0712	0.1674
82 Oberholzer et al. (2024)	Study 1 (Switzerland, lab)	2.0254	5.9679	0.0230	0.0594
83 Oberholzer et al. (2024)	Study 2 (US, Prolific)	2.5519	8.1319	0.0287	0.0762
84 Oprea (2024)	Main (online MPL)	-0.0435	4.4776	0.0132	0.1870
85 Oprea (2024)	Student lab	0.5140	2.0997	0.0274	0.0927
86 Oprea (2024)	BDM elicitation	0.2653	2.2305	0.0202	0.0960
87 Oprea (2024)	4-box display	0.0834	2.7929	0.0110	0.1044
88 Oprea (2024)	Error robustness	0.3437	3.9677	0.0264	0.1760
89 Vieider et al. (2019)	Vietnam, farmers	-2.6518	9.3645	-0.0493	0.4178
90 Vieider et al. (2019)	Vietnam, students	-1.3080	3.9879	-0.0425	0.1870

NOTES. This table reports the estimated location ($\hat{\tau}$) and scale ($\hat{\sigma}$) parameters for each of the 90 treatments, estimated using specification (iv) of the Baseline model.

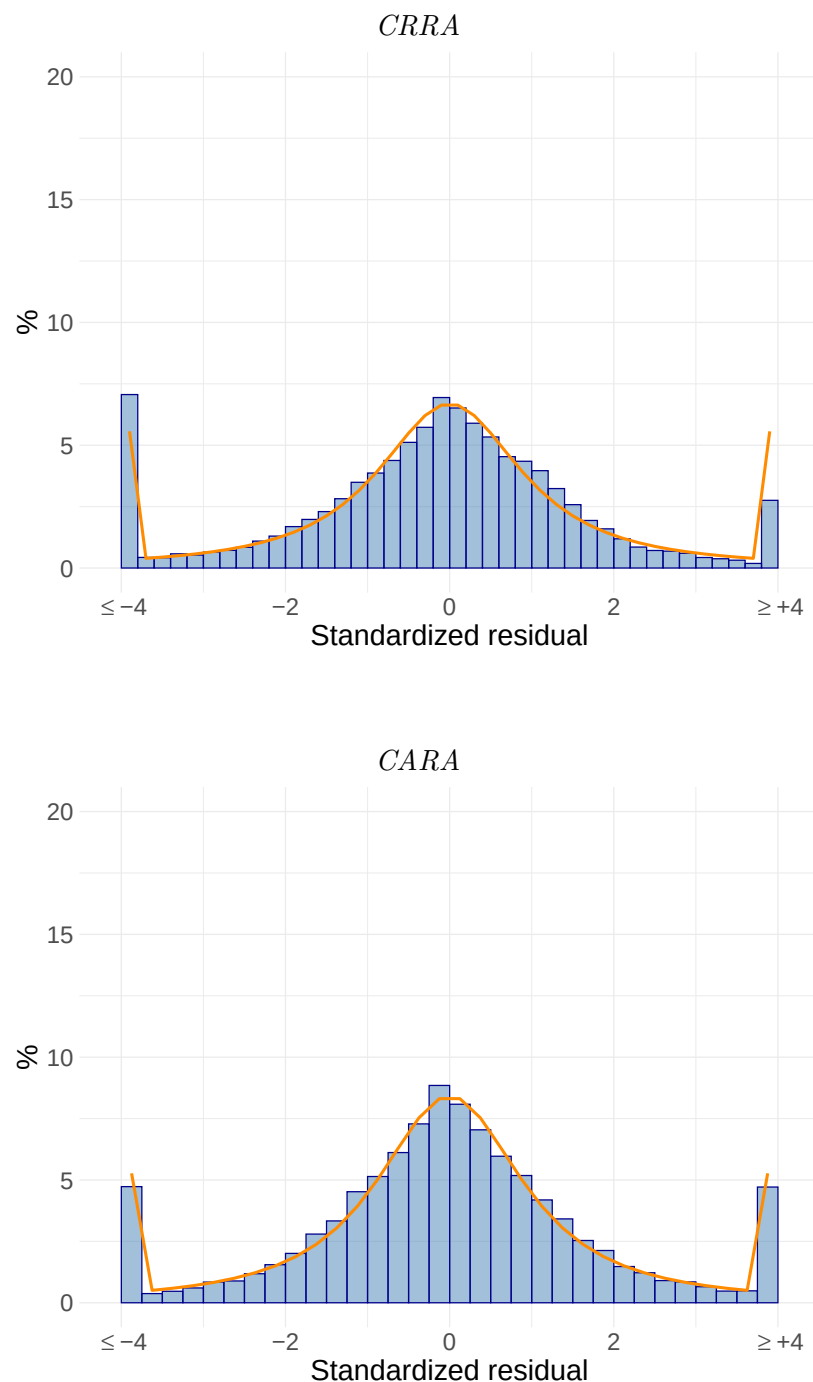
APPENDIX B. STUDENT- t DISTRIBUTION

FIGURE B1. Baseline with a Student- t Distribution: Distribution of Standardized Risk Aversion Coefficients



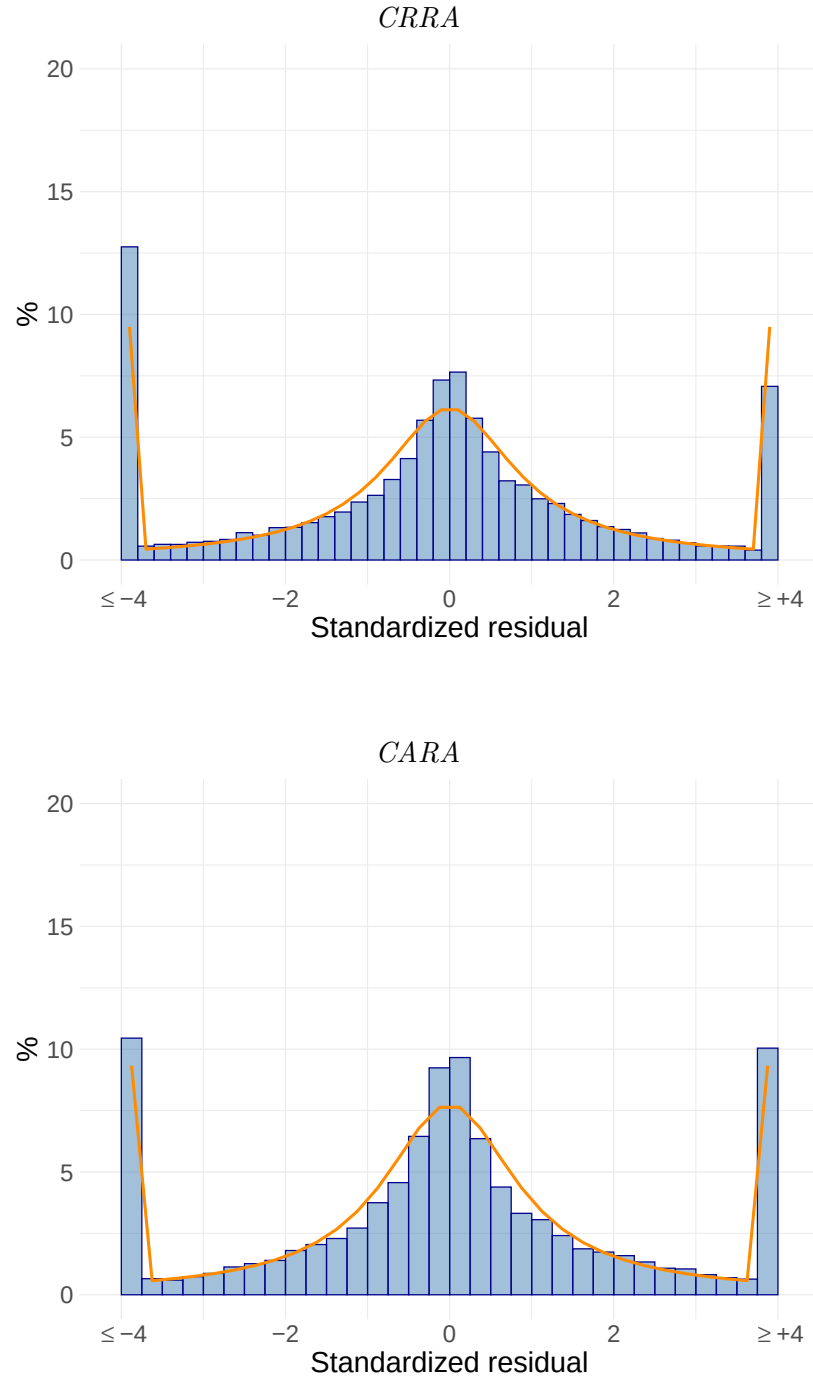
NOTES. The figure shows histograms of the risk aversion coefficients standardized using the estimated location and scale parameters of the corresponding treatment with a Student- t distribution. The top row corresponds to the CRRA utility function, and the bottom row to the CARA utility function. In each panel, the PDF of the Student- t distribution using $\hat{\nu}$ is shown in orange.

FIGURE B2. Lottery-Specific Noise with a Student- t Distribution:
Distribution of Standardized Risk Aversion Coefficients



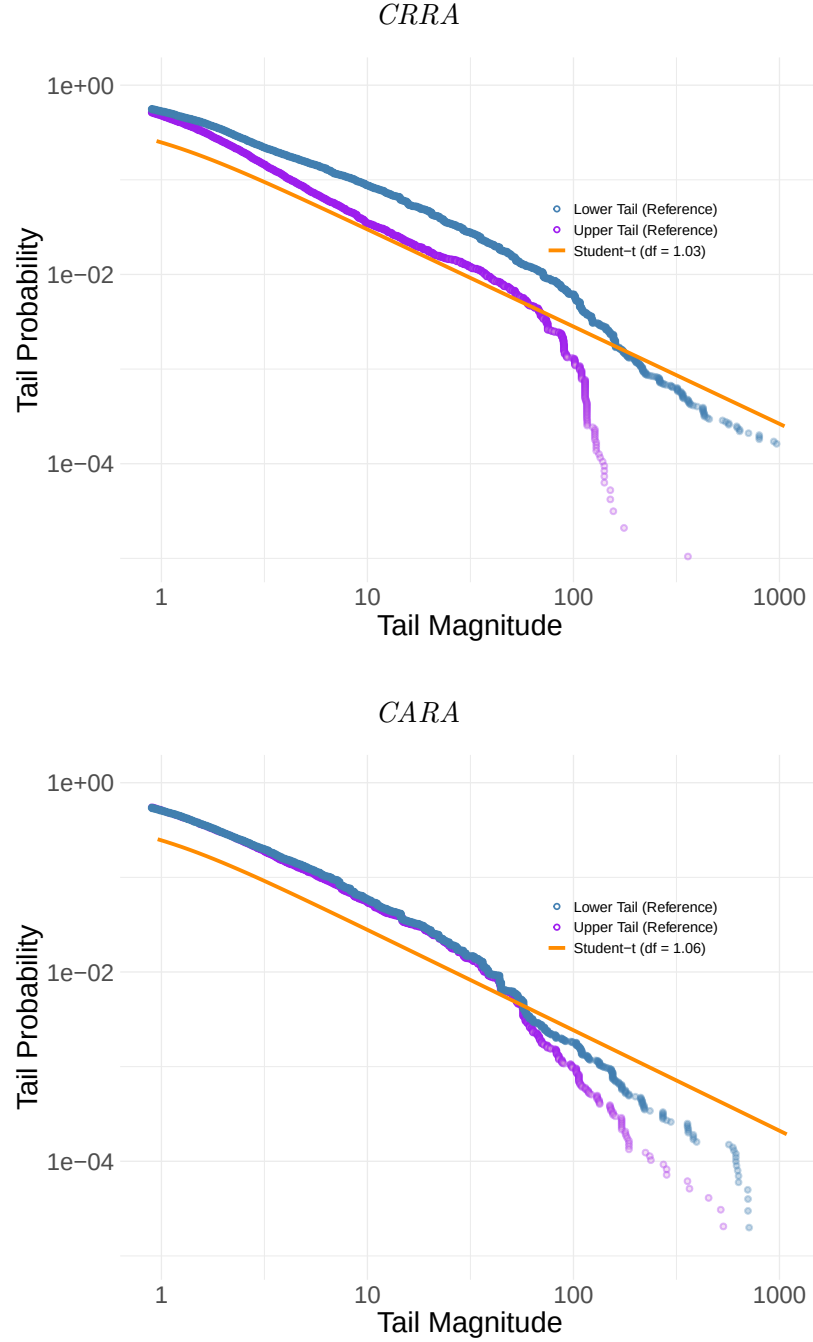
NOTES. The figure shows histograms of the risk aversion coefficients standardized using the estimated location and scale parameters from the saturated specification with a Student- t distribution. The top row corresponds to the CRRA utility function, and the bottom row to the CARA utility function. In each panel, the PDF of the Student- t distribution using $\hat{\nu}$ is shown in orange.

FIGURE B3. Lottery-Specific (Mean) Risk with a Student- t Distribution: Distribution of Standardized Risk Aversion Coefficients



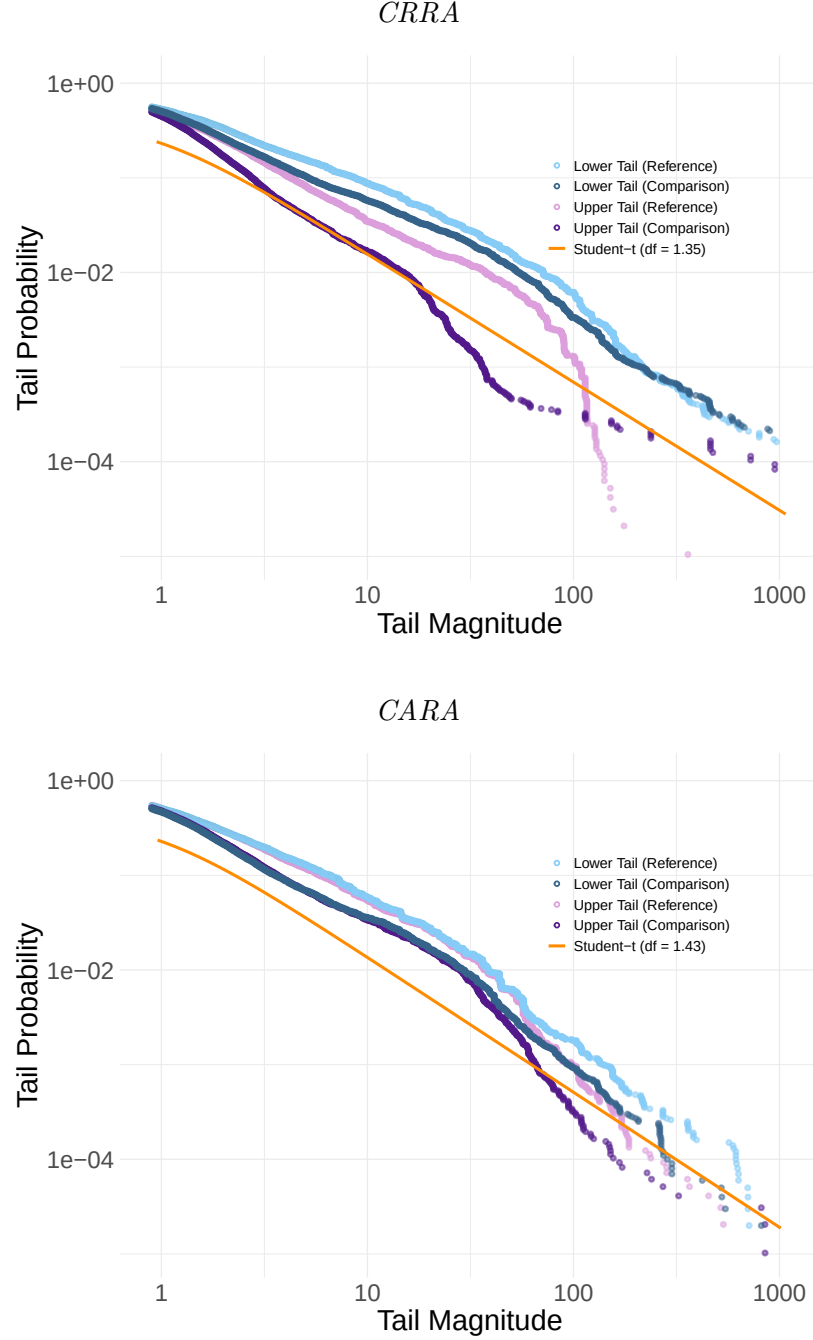
NOTES. The figure shows histograms of the risk aversion coefficients standardized using the estimated location and scale parameters from the saturated specification with a Student- t distribution. The top row corresponds to the CRRA utility function, and the bottom row to the CARA utility function. In each panel, the PDF of the Student- t distribution using $\hat{\nu}$ is shown in orange.

FIGURE B4. Baseline with a Student- t Distribution: Log-Log Plots of the Tails of the Empirical Distribution of Standardized Risk Aversion Coefficients



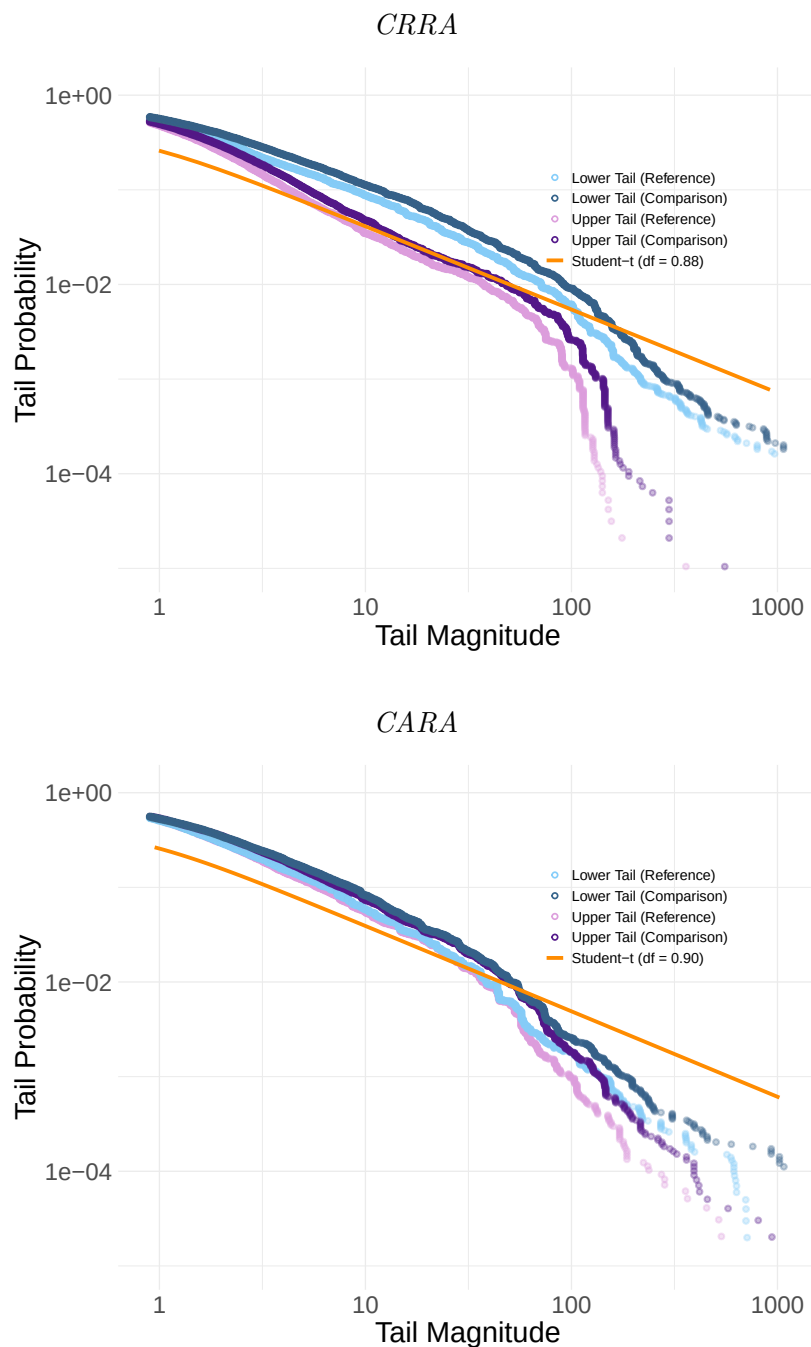
NOTES. The figure shows log-log plots of the tails of the empirical distribution of the risk aversion coefficients standardized using the estimated location and scale parameters of the corresponding treatment with a Student- t distribution. The top row corresponds to the CRRA utility function, and the bottom row to the CARA utility function. In each panel, the tail of the Student- t PDF using $\hat{\nu}$ is shown in orange.

FIGURE B5. Lottery-Specific Noise vs Baseline with a Student- t Distribution: Log-Log Plots of the Tails of the Empirical Distribution of Standardized Risk Aversion Coefficients



NOTES. The figure shows log-log plots of the tails of the empirical distribution of the risk aversion coefficients standardized using the estimated location and scale parameters from the saturated specification with a Student- t distribution. The top row corresponds to the CRRA utility function, and the bottom row to the CARA utility function. Each panel plots two sets of tails: the series labeled *Reference* in the legend correspond to the Baseline specification (lighter shades), and those labeled *Comparison* to the specification with lottery-specific noise (darker shades). In each panel, the tail of the Student- t PDF using $\hat{\nu}$ is shown in orange.

FIGURE B6. Lottery-Specific (Mean) Risk vs Baseline with a Student- t Distribution: Log-Log Plots of the Tails of the Empirical Distribution of Standardized Risk Aversion Coefficients



NOTES. The figure shows log-log plots of the tails of the empirical distribution of the risk aversion coefficients standardized using the estimated location and scale parameters from the saturated specification with a Student- t distribution. The top row corresponds to the CRRA utility function, and the bottom row to the CARA utility function. Each panel plots two sets of tails: the series labeled *Reference* in the legend correspond to the Baseline specification (lighter shades), and those labeled *Comparison* to the specification with lottery-specific (mean) risk (darker shades). In each panel, the tail of the Student- t PDF using $\hat{\nu}$ is shown in orange.

APPENDIX C. ADDITIONAL CROSS-VALIDATION ANALYSIS

The main text reports the leave-one-lottery-out (LOLO) exercise, which removes a single lottery at a time while keeping its study, treatment, and subjects in the training sample. Here we complement it with two more demanding exercises that hold out entire aggregates: leave-one-study-out (LOSO), which removes every observation of one study (15 folds), and leave-one-treatment-out (LOTO), which removes every observation of one experimental treatment (90 folds). Because all subjects in the held-out unit are removed, the model must predict for a study or treatment it has never seen while limiting potential interpolation from information about similar lotteries in the dataset. Consequently, both provide more conservative tests of out-of-sample validation than LOLO.

Holding out an entire unit raises a practical difficulty: the location and scale of a study or treatment are governed by fixed effects that the training data cannot identify. We resolve it with a *calibrated* fixed-effect design that separates the parameters we expect to transfer across units from those that are inherently unit-specific. In a first step, using specification (iv), the main parameters of interest (the lottery-characteristic coefficients β and, under the Student- t distribution, the degrees of freedom ν) are estimated; these parameters are then carried over to the held-out unit. In a second step, the subjects of the held-out unit are split into calibration and validation subsamples and we estimate the held-out unit’s location and scale intercepts, holding β and ν fixed, on the calibration subsample. In a third step, the disjoint validation subsample is scored. The split is a cross-fitting partition of the held-out unit’s subjects (involving 5 folds), so every subject is scored exactly once and only the unit’s location and scale intercepts are re-estimated on each calibration subsample.

We evaluate predictions with the logarithmic scoring rule, where each held-out observation contributes the log predictive density that the model assigns to the choice actually made, so larger values reward sharper and better-calibrated predictions.¹ For every specification, we report the observation-weighted mean difference in log-score relative to the Baseline model, denoted as Δ Log-Score, which is formed choice by choice on the same held-out observations. A positive value of Δ Log-Score means the specification places a higher predictive density on the realized choices than the already-rich Baseline does. Confidence intervals are clustered at the subject level. To aid interpretation, we also show two monotone transformations of this measure. The first is the Density Gain = $\exp(\Delta) - 1$, which expresses the improvement as a proportional increase in predictive density per observation. The second is the Unit Win-Rate, which shows the share of held-out units (studies or treatments) with a positive mean gain.

Table C1 reports the held-out gains under both the logistic distribution used in the main analysis and the Student- t distribution used for robustness. The gains mirror the LOLO

¹For more details, see Gneiting, T., and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.

exercise. Across both families and both channels, the specifications improve on the Baseline: observation-weighted Δ log-scores are positive throughout, with confidence intervals that exclude zero in all but two cases, density gains between roughly 0.3% and 15.6%, and per-unit win-rates between 52% and 89%. Since the Student- t distribution better captures the tails, the larger gains associated with the noise (scale) channel are attenuated.

Three checks support these conclusions. First, every fit converged; no treatment had to be dropped due to insufficient subjects for both calibration and validation. Second, re-running the inner cross-fitting with three or ten folds, rather than five, leaves the gains essentially unchanged, so they are not an artifact of a particular calibration/validation split. Third, the predictive distribution is well calibrated particularly under the Student- t distribution.

TABLE C1. Calibrated Fixed-Effect Cross-Validation: Logistic and Student- t Distributions

Family	Specification	Δ Log-Score [95% CI]	Density Gain	Unit Win-Rate
Upper Panel: Logistic Distribution				
<i>Panel A. Leave-One-Study-Out (LOSO), 15 studies</i>				
CRRA	Noise+ Z_L	+0.006 [−0.016, +0.029]	+0.6%	73%
	(Mean) Risk+ Z_L	+0.008 [+0.002, +0.014]	+0.8%	60%
CARA	Noise+ Z_L	+0.036 [+0.019, +0.053]	+3.7%	73%
	(Mean) Risk+ Z_L	+0.003 [−0.000, +0.007]	+0.3%	60%
<i>Panel B. Leave-One-Treatment-Out (LOTO), 90 treatments</i>				
CRRA	Noise+ Z_L	+0.095 [+0.073, +0.118]	+10.0%	62%
	(Mean) Risk+ Z_L	+0.024 [+0.020, +0.028]	+2.4%	52%
CARA	Noise+ Z_L	+0.145 [+0.129, +0.161]	+15.6%	84%
	(Mean) Risk+ Z_L	+0.012 [+0.009, +0.015]	+1.2%	73%
Lower Panel: Student-t Distribution				
<i>Panel A. Leave-One-Study-Out (LOSO), 15 studies</i>				
CRRA	Noise+ Z_L	+0.022 [+0.018, +0.027]	+2.2%	87%
	(Mean) Risk+ Z_L	+0.047 [+0.042, +0.051]	+4.8%	87%
CARA	Noise+ Z_L	+0.046 [+0.040, +0.052]	+4.7%	87%
	(Mean) Risk+ Z_L	+0.008 [+0.004, +0.012]	+0.8%	67%
<i>Panel B. Leave-One-Treatment-Out (LOTO), 90 treatments</i>				
CRRA	Noise+ Z_L	+0.048 [+0.044, +0.052]	+4.9%	72%
	(Mean) Risk+ Z_L	+0.053 [+0.050, +0.057]	+5.5%	74%
CARA	Noise+ Z_L	+0.081 [+0.076, +0.087]	+8.5%	89%
	(Mean) Risk+ Z_L	+0.022 [+0.019, +0.026]	+2.3%	86%

The upper (lower) panel reports on the logistic (Student- t) distribution. Held-out Δ log-score relative to the Baseline model under the calibrated fixed-effect design. In each case, Panel A reports on LOSO and Panel B on LOTO. Δ Log-Score is the observation-weighted mean gain over Baseline, with 95% confidence intervals from subject-level clustering; Density Gain is $\exp(\Delta) - 1$, the proportional increase in predictive density per observation; Unit Win-Rate is the share of held-out units with a positive mean gain.