risks

MDPI

Article

When Should Demographics Enter the Prior? Conditional Exchangeability in Bayesian Estimation of Risk Preferences

Xiaoxue Sherry Gao [1,*] and Glenn W. Harrison [2,3]

[1] Department of Resource Economics, University of Massachusetts Amherst, Amherst, MA 01003, USA
[2] Maurice R. Greenberg School of Risk Sciences and Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University, Atlanta, GA 30303, USA; gharrison@gsu.edu
[3] School of Economics, University of Cape Town, Cape Town 7701, South Africa
* Correspondence: xiaoxuegao@umass.edu

Abstract

Hierarchical Bayesian models for the estimation of individual risk preferences typically assume that subjects are unconditionally exchangeable: prior to seeing any data, every subject is treated as a draw from a common population distribution. We examine the consequences of relaxing that assumption to one of conditional exchangeability, where the hyper-parameters of the population distribution are allowed to depend on observable covariates. Using simulated data, where the true parameters are known, we compare the recovery properties of unconditionally and conditionally exchangeable specifications across both Expected Utility and Rank-Dependent Utility models. A parsimonious conditionally exchangeable specification recovers individual risk preferences well when the data-generating process actually contains demographic structure, and imposes only a modest penalty when it does not, but overly rich covariate specifications can substantially erode the borrowing-of-strength that makes hierarchical models attractive in the first place. The trade-off is sharp, but navigable, and is of direct relevance to applied work that uses estimated risk preferences as inputs to normative welfare analysis or as controls for other inferences.

Keywords: risk preferences; Bayesian estimation; hierarchical models; exchangeability; heterogeneity

JEL Classification: D81; C11; C51

Individual risk preferences are central to a remarkable range of descriptive and normative inferences in economics. Knowing them is necessary to evaluate insurance and savings decisions, to predict portfolio composition, to interpret behavior in auctions, to identify time preferences from intertemporal choices over risky outcomes, and to assess whether observed choices generate welfare gains or losses. Across all of these uses, what is needed is a defensible estimate of the risk preferences *of each individual*, not just of an unspecified representative agent. And it is now well understood that risk preferences are heterogeneous across individuals in ways that matter quantitatively as well as qualitatively.

Estimating risk preferences at the individual level is hard. The data one can obtain from any one subject are typically limited: time is scarce, attention is finite, and the experimental tasks needed to identify utility curvature, probability weighting, and stochastic noise all compete for the subject's effort. Classical maximum likelihood (ML) estimation at the individual level routinely leads to imprecise, occasionally absurd, and sometimes simply non-existent estimates. Bayesian hierarchical models (BHM) offer a disciplined response.

Check for updates

Academic Editor: Zbigniew Palmowski

Received: 22 May 2026
Revised: 6 July 2026
Accepted: 28 July 2026
Published: 4 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

By treating each subject's parameters as draws from a common population distribution, they let information from the rest of the sample inform inference about any single subject. When that subject's own data are uninformative, the population prior does the heavy lifting; when they are informative, the data dominate. Gao et al. (2023) document the practical gains from this approach in the context of behavioral welfare evaluations of insurance choices that depend on the estimated risk preferences of individuals. The literature on hierarchical Bayesian estimation of preferences in marketing, psychology, and behavioral economics is by now considerable, going back at least to Rossi and Allenby (1993), Allenby and Rossi (1999), Nilsson et al. (2011), and the textbook treatments of Gelman et al. (2013), Johnson et al. (2012) and Kruschke (2015).

Most of that literature, including our own Gao et al. (2023), rests on an assumption of *unconditional exchangeability*: before seeing any data, the analyst regards the subjects as interchangeable draws from a common distribution. That is a strong assumption, and a deliberate one. It is what makes the hierarchical model *hierarchical*, in the sense that it permits pooling. The price of that pooling is that the model knows nothing, *a priori*, about what distinguishes one subject from another.

In practice, of course, the analyst usually *does* know something about each subject. Gender, age, education, prior insurance experience, income, and many other observables are routinely recorded. A natural question is whether allowing the hyper-parameters of the population distribution to depend on these observables, or *conditional exchangeability*, improves the recovery of individual risk preferences. The intuition for an improvement is straightforward. If women in our subject pool really are, on average, more risk averse than men, then knowing a subject's gender is genuinely informative about her risk preferences, and a model that uses that information should produce appropriately sharper posteriors. The intuition for a cost is equally straightforward, and is the point we want to develop. Each additional covariate carves the population into smaller cells. In the limit, with enough covariates and interactions, each subject ends up in a cell of her own, and the borrowing-of-strength that made the hierarchical model useful in the first place evaporates.

We examine that trade-off directly. Using simulated data, generated under known data-generating processes, we evaluate how well alternative BHM specifications recover individual risk preferences. The specifications differ in whether they assume unconditional exchangeability, conditional exchangeability with a handful of demographic covariates, or no exchangeability at all (estimation one subject at a time). We do this for both Expected Utility Theory (EUT) and Rank-Dependent Utility (RDU) data-generating processes, and then turn to observed choices from a laboratory experiment and ask, with appropriate Bayesian caution, which specification best describes the data and how the inferences about individual risk preferences differ across specifications.

Our main findings are three. First, when demographic structure is genuinely present in the data, a conditionally exchangeable model with a few covariates recovers individual risk preferences more accurately than its unconditionally exchangeable counterpart. Second, when demographic structure is *absent*, the penalty for, nonetheless, including a few covariates is small: the posterior simply concludes that the covariates do not matter, and the individual-level estimates are very close to those of the unconditional model. Third, and most usefully for applied work, the gains from conditioning are bounded. With too many covariates, or with interactions that effectively put each subject in a unique cell, the conditioning destroys the very pooling that the hierarchical structure is meant to deliver. The practical guidance that emerges is to condition parsimoniously, on covariates with a credible *a priori* connection to risk preferences, and to use Bayes factors as one input, alongside theoretical priors and the inference at hand, when evaluating whether the conditioning has earned its keep.

Section 1 sets out the exchangeability assumption: what it means, where it comes from, and why it is the foundation of any hierarchical Bayesian model. It then explains conditional exchangeability, situates it within the broader literature (including its more recent prominence in causal inference), and describes how we implement it with covariates that enter linearly into the hyper-parameters of the population distribution. The discussion closes with the trade-off between informativeness and the loss of pooling. Section 2 describes the experimental data we use both as the basis for our simulations and for our estimates from observed choices. Section 3 presents the EUT specifications, the simulation evidence on parameter recovery and sensitivity to mis-specification, and the estimates from observed data. Section 4 extends the analysis to RDU. Section 5 discusses the limitations of our analysis and the extensions that they invite. Section 6 concludes the paper.

1. The Exchangeability Assumption

1.1. Unconditional Exchangeability

A sequence of random variables is *exchangeable* if the joint distribution of any finite subset of them is invariant to permutations of the indices. Loosely: the labels on the observations do not matter. If we relabel subject 1 as subject 17, and subject 17 as subject 1, our beliefs about the joint distribution of their parameters should be unchanged. The formal study of exchangeability goes back to de Finetti (1937), whose representation theorem established that any infinite exchangeable sequence of random variables can be represented as conditionally independent and identically distributed, given some unobserved parameter drawn from a prior distribution. In other words, exchangeability *is* what justifies the language of "drawing parameters from a population distribution" in the first place. Without exchangeability, there is no coherent Bayesian sense in which the subjects in a sample tell us anything about each other.[1]

In the context of a hierarchical Bayesian model of risk preferences, unconditional exchangeability means that, before observing any choices, the analyst regards every subject in the sample as a draw from a common distribution. If r_i denotes the utility curvature parameter of subject i, the model writes

$$r_i \mid m_r,\, \sigma_r^2 \;\sim N\left(m_r,\, \sigma_r^2\right),$$

and places hyper-priors over m_r and σ_r^2. The same population parameters (m_r, σ_r^2) govern every subject. The names of the subjects are not arguments of the distribution. The hierarchical structure does the rest of the work: when the data for any one subject are uninformative, the posterior for that subject's r_i is pulled toward the population mean. When the data are informative, the posterior moves away from the population mean. This "shrinkage" toward a common centre is, in classical language, what makes hierarchical estimates a regularized improvement on independent subject-by-subject estimates.[2]

The strength of unconditional exchangeability is that it is robust. It commits the analyst to nothing beyond the claim that the subjects are, *a priori*, interchangeable. The weakness is exactly the same thing. If the subjects in fact differ systematically along observable dimensions, i.e., if women are systematically more risk averse than men in this sample, the model is required to ignore that information at the prior stage. The information returns through the data, but only one subject at a time, and the pooling that takes place is pooling toward a single population mean that may not be the right center for any sub-group.

1.2. Conditional Exchangeability

Conditional exchangeability weakens the requirement such that subjects are exchangeable only *after* conditioning on a vector of observables. Two subjects with the same observed

covariates are interchangeable; two subjects with different covariates need not be. Formally, we write

$$r_i \mid m_r,\ \beta,\ \sigma_r^2,\ X_i \ \sim N\left(m_r + \beta \cdot X_i,\ \sigma_r^2\right),$$

where X_i is the vector of observed covariates for subject i and β is a vector of hyper-coefficients to be estimated. The population mean is now allowed to vary with observables, while the population variance σ_r^2 continues to capture residual heterogeneity among subjects with the same covariate values.[3] That is, of course, just a regression structure embedded in the prior, and that is precisely what it is meant to be. The parameters (m_r, β, σ_r^2) are themselves estimated from the data, jointly with the individual r_i.

The intellectual roots of conditional exchangeability run deep, and in directions that may surprise economists. Early formal treatments are in Diaconis and Freedman (1980), and the idea has been a workhorse in Bayesian hierarchical modeling since at least Lindley and Smith (1972). It has acquired renewed prominence in the modern causal inference literature, where conditional exchangeability of treated and untreated units, given observed covariates, is the central identifying assumption behind the entire family of selection-on-observables estimators, such as propensity-score methods, doubly robust estimators, matching, regression adjustment, and the recent generation of machine-learning-based causal estimators discussed in Imbens and Rubin (2015), Hernán and Robins (2020) and Athey and Imbens (2017). The terminology there is sometimes "ignorability" or "unconfoundedness", but the substantive content is the same: conditional on observables, the labels stop carrying information. Our use of the assumption is in a different setting, Bayesian estimation of individual risk preferences from experimental data, not estimation of treatment effects from observational data, but the underlying logic is identical.

1.3. Implementation with Linear Covariates

We implement conditional exchangeability by allowing the hyper-mean of each individual-level parameter to depend linearly on a small set of demographic covariates. For our subject pool we use five binary covariates: whether the subject is female, whether the subject is Black, whether the subject is a business major, whether the subject has a high GPA, and whether the subject owns any insurance. These are the covariates that have been used in related work on this population, and they are the covariates for which a credible *a priori* case can be made that they could move risk preferences in our sample. For the utility curvature parameter, we therefore write

$$r_i \ \sim N\left(m_r + \beta \cdot X_i,\ \sigma_r^2\right),$$

with diffuse hyper-priors on each component of β. For the Fechner noise parameter μ_i, and (under RDU) for the two probability-weighting parameters η_i and φ_i, we use analogous specifications on the log scale, since these parameters are constrained to be positive. The full specifications are given in the next two sections, where we describe the EUT and RDU models.

1.4. The Trade-Off

The promise of conditional exchangeability is sharper inference. If the covariates really do shift the population mean, then knowing a subject's covariates tells us where, on average, her parameter sits, and the posterior for that subject benefits from a better-centered prior. The threat is more subtle. With one binary covariate, the model effectively divides the sample into two pools and shrinks each subject's estimate toward the mean of her pool. That is generally helpful. With five binary covariates, the model is, implicitly, defining up to $2^5 = 32$ cells; with interactions, far more. As the cells become smaller, the population

mean for each cell is estimated from fewer subjects, and the gain from pooling within a cell falls. In the limit, every subject is in her own cell, the conditioning has eliminated all pooling, and we are back to estimating each subject's parameters from her own data alone, which is the worst case the hierarchical model was designed to avoid.

The trade-off is therefore not between conditional and unconditional exchangeability in the abstract, but between the right amount of conditioning and too much. The simulations in the next sections are designed to characterize that trade-off concretely. We deliberately consider data-generating processes that differ in whether demographic structure is present, and specifications that differ in whether and how it is modeled. The combination tells us, for our setting of risk preferences, what conditioning helps, what conditioning hurts, and what conditioning makes no difference.

Two practical points are worth flagging in advance. The first is that Bayes factors and posterior model probabilities provide a natural way to *evaluate* how well alternative amounts of conditioning describe the observed choices, and we use them in that spirit throughout.[4] We deliberately say "evaluate" rather than "select". Model-selection language treats a Bayes factor as a verdict that eliminates one model in favor of another for all purposes, which is more than the metric can carry. The Bayes factor is one summary of in-sample predictive consistency under the priors actually used, and that is exactly the question we want to ask of our specifications. But it is not a verdict about which utility model should be carried forward into, say, a downstream welfare evaluation, where the structural form of the model is constrained by the inference being made rather than by a horse-race statistic. The second point is that the trade-off described here is not unique to risk preferences. Anyone who has used regression with covariates to estimate group-level means, or used a multilevel model with random intercepts that depend on observables, has implicitly been navigating the same trade-off. The novelty of what we do is to confront it head-on, in the specific context of individual-level structural parameters that are themselves inputs to downstream economic inference.

2. Experimental Data

We have N = 73 subjects each making T = 60 lottery choices. The subjects are recruited from the undergraduate student population of Georgia State University. The 60 pairs of lotteries are each in the gain domain: for any k = 1, . . ., K, outcome $x_k \geq 0$ and subjects receive an endowment e of \$0. We use subscript i = 1, 2, . . ., 73 to index the subject, t = 1, 2, . . ., 60 to index the tth lottery pair, and $y_{i,t} = 1$ to indicate that subject i chose the lottery on the left in pair t. The labels "left" and "right" are purely positional and have no economic content; they refer only to where the option appeared on the screen. We collect all decisions of one subject in $y_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,60})$ and all decisions of all subjects in $y = (y_1, y_2, \ldots, y_{73})$.

Each subject was asked to make choices over pairs of lotteries designed to provide evidence of risk aversion as well as the tendency to make decisions consistently with EUT or RDU models. The battery of lottery choices combines designs from Loomes and Sugden (1998), which vary the gradient of the EUT-consistent indifference curves within a Marschak–Machina triangle so as to provide tests of EUT with adequate power against a range of risk attitudes; designs from Harrison et al. (2015) for testing the Reduction of Compound Lotteries axiom; and a series of lotteries that are actuarially equivalent to insurance choices studied by Harrison et al. (2025). Each subject faced an individually randomized sequence of choices from this battery. The typical screen interface, full instructions, and the complete list of lottery parameters are documented in Appendix A.

Subjects were paid for one of their choices. Prizes ranged from \$0 to \$70 across the lotteries, with average expected earnings of \$27.80 conditional on observed choices. Each subject also received a \$7 participation fee.

In addition to the choice data, we observe a vector of demographic characteristics for each subject. These characteristics provide the covariates X_i that enter the conditionally exchangeable specifications. Summary statistics are reported in Appendix B.

3. Expected Utility Theory

We use the EUT model to develop our analysis of conditional versus unconditional exchangeability in a setting that is simple enough to be transparent and rich enough to be useful. The first subsection lays out the model and the six specifications we compare. The second subsection uses simulated data, where the true parameters are known, to evaluate parameter recovery. The third subsection uses observed choices.

3.1. Model and Specifications

Individuals are assumed to integrate lottery prizes perfectly with their endowments and to evaluate prizes through a Constant Relative Risk Aversion (CRRA) utility function $u(e,\ x_k) = (e + x_k)^{1\ -\ r}\ /\ (1\ -\ r)$ for any k = 1, …, K. In our battery, K = 4. A lottery is evaluated by the probability-weighted sum of utilities of prizes, and the latent index for choice t by subject i is the difference between the EU of the left and right lotteries scaled by a Fechner noise parameter μ_i and a contextual utility term that normalizes utilities of prizes between 0 and 1, following Wilcox (2011). The probability of choosing the left lottery is then a logistic function of this latent index. Collecting the data-generating process in compact form, the two individual-level parameters are the utility curvature r_i and the Fechner noise μ_i.

We consider six econometric specifications, indexed I through VI, that differ in the degree of heterogeneity they allow for these two parameters. All are nested in the most general specification (VI), which allows for heterogeneity at both the demographic and the individual level. We describe VI first and treat the others as restrictions of it.

Specification VI is fully conditionally exchangeable with individual heterogeneity. The individual CRRA coefficient is drawn from the normal distribution specified in Section 1.3, and μ_i drawn from a log-normal distribution with a common mean and variance. The hyper-priors on m_r, the components of β, and $m_{ln\ \mu}$ are diffuse N(0, 100), and the hyper-priors on σ_r and $\sigma_{ln\ \mu}$ are inverse-Gamma.[5]

The remaining specifications are restrictions of VI obtained by setting one or more of its features to zero. Specification V drops the demographic component ($\beta = 0$) but retains individual heterogeneity in both r_i and μ_i; this is the unconditionally exchangeable hierarchical model in the sense of Section 1.1. Specification IV retains conditional exchangeability for r_i but eliminates within-cell heterogeneity ($\sigma_r^2 = 0$), so subjects with identical covariates are required to have identical r_i. Specification III is the representative-agent model with no demographic structure and no individual heterogeneity in r_i, but allows for heterogeneity in the noise parameter μ_i. Specifications II and I correspond to IV and III, respectively, but additionally impose homogeneity of μ_i across subjects. We estimate all models in Stata using a combination of the Metropolis–Hastings algorithm and the Gibbs sampler, with 10,000 MCMC draws after a burn-in of 2500 in all reported estimations. The Stata template that defines the likelihood can be found in Online Appendix A of Gao et al. (2023).

3.2. Parameter Recovery

To evaluate the recovery properties of the six specifications under known data-generating processes, we simulate six datasets, indexed I through VI, that correspond

to the same six structures: each subject's true parameters are drawn from the structure that the corresponding specification assumes. The simulation parameters used to generate each dataset are summarized in Table 1. These parameters are displayed in italics in Table 1 and all the estimation tables to follow to differentiate them from estimated values. For each simulated dataset, we use the observed lottery characteristics and demographic covariates of our 73 subjects, but draw the risk preference parameters from the specified distribution and then simulate the binary choices. This gives us simulated data with the same structure as the observed data, but with parameters we control.

Table 1. Parameters in simulated datasets.

Sim.	Constant (r)	Female (r)	Black (r)	BusinessMajor (r)	HighGPA (r)	Insured (r)	Std. Dev. (r)	Mean (μ)	Std. Dev. (μ)
I	*0.5*	-	-	-	-	-	-	*0.25*	-
II	*0.4*	*0.2*	*−0.2*	*−0.5*	*0.1*	*0.3*	-	*0.25*	-
III	*0.5*	-	-	-	-	-	-	*0.23*	*0.07*
IV	*0.4*	*0.2*	*−0.2*	*−0.5*	*0.1*	*0.3*	-	*0.23*	*0.07*
V	*0.5*	-	-	-	-	-	*0.2*	*0.23*	*0.07*
VI	*0.4*	*0.2*	*−0.2*	*−0.5*	*0.1*	*0.3*	*0.2*	*0.23*	*0.07*

We first estimate each simulated dataset under its correct specification, to evaluate whether the BHM can recover what we put in. All estimations converge with moderate to high efficiency; see Appendix C for a full diagnostic report. Estimates of the population-level parameters are reported in Table 2. The true values used in the simulations are inside the 89% credible intervals for almost every parameter in every specification.[6] The only exception is the standard deviation σ_r in specification VI, where the posterior is somewhat tighter than the true value; this happens because, under specification VI, σ_r measures only the residual heterogeneity around the demographic-conditional mean, and the model assigns some of that residual variance to the demographic component. The mis-identification is statistical, not substantive: as we show below, the posterior conclusions about individual r_i from specification VI on observed data are unchanged whether or not σ_r is recovered exactly.

Recovery at the individual level is the more interesting test. For specifications that allow individual heterogeneity in r_i, we compare the posterior mean of each subject's r_i to the true value used in the simulation. Figure 1 shows the resulting scatter plots. For simulated datasets II and IV, where there is no within-cell heterogeneity, the recovery is essentially perfect: the points lie on the 45° line and the correlation between true and estimated values is close to 1. For dataset VI, where both demographic and individual heterogeneity are present, the correlation is 0.94 and the cloud is tight. For dataset V, where there is individual heterogeneity but no demographic structure, the correlation is 0.77 and the points at the tails are pulled toward the center by the shrinkage prior, exactly as the hierarchical model is meant to do. Across all four datasets, the 89% credible intervals contain the true value for the vast majority of subjects (Table A6). The BHM recovers what we put in.

We now turn to the question of what happens when we estimate dataset VI, the most demanding data-generating process, under incorrect specifications. The exercise is designed to make the trade-off described above concrete. Dataset VI has both demographic structure ($\beta \neq 0$) and within-cell individual heterogeneity ($\sigma_r^2 > 0$). Specification V drops the demographic structure but retains within-cell heterogeneity. Specification IV drops within-cell heterogeneity but retains the demographic structure. Specification III drops both. Specifications II and I additionally drop heterogeneity in μ_i. Estimates under each of

these mis-specifications are reported in Tables 3 and A7 of Appendix B; population-level estimates are in the main text and individual-level estimates are deferred to Appendix B.

Table 2. Expected utility models using simulated choices under correct specifications.

Group	Parameter	Spec	True	Mean	SD	89% CI Low	89% CI High
r	Constant	***I***	*0.5*	0.503	0.017	0.479	0.532
		II	*0.4*	0.351	0.055	0.258	0.432
		III	*0.5*	0.496	0.015	0.472	0.518
		IV	*0.4*	0.377	0.043	0.308	0.446
		V	*0.5*	0.493	0.025	0.453	0.533
		VI	*0.4*	0.359	0.071	0.234	0.463
r	Female	***II***	*0.2*	0.214	0.036	0.156	0.270
		IV	*0.2*	0.193	0.032	0.146	0.246
		VI	*0.2*	0.191	0.056	0.091	0.271
r	Black	***II***	*−0.2*	−0.206	0.031	−0.257	−0.159
		IV	*−0.2*	−0.205	0.027	−0.248	−0.163
		VI	*−0.2*	−0.181	0.058	−0.272	−0.089
r	BusinessMajor	***II***	*−0.5*	−0.457	0.043	−0.524	−0.390
		IV	*−0.5*	−0.480	0.042	−0.546	−0.413
		VI	*−0.5*	−0.488	0.064	−0.595	−0.391
r	HighGPA	***II***	*0.1*	0.093	0.034	0.033	0.144
		IV	*0.1*	0.088	0.031	0.041	0.138
		VI	*0.1*	0.117	0.059	0.024	0.217
r	Insured	***II***	*0.3*	0.327	0.033	0.276	0.379
		IV	*0.3*	0.324	0.028	0.277	0.364
		VI	*0.3*	0.304	0.061	0.207	0.405
r	Std. Dev.	***V***	*0.2*	0.169	0.022	0.135	0.204
		VI	*0.2*	0.098 *	0.014	0.076	0.119
μ	Mean	***I***	*0.25*	0.260	0.014	0.239	0.281
		II	*0.25*	0.242	0.011	0.224	0.259
		III	*0.23*	0.240	0.018	0.212	0.268
		IV	*0.23*	0.221	0.014	0.199	0.242
		V	*0.23*	0.230	0.017	0.203	0.254
		VI	*0.23*	0.210	0.016	0.184	0.234
μ	Std. Dev.	***III***	*0.07*	0.082	0.026	0.044	0.125
		IV	*0.07*	0.061	0.019	0.031	0.089
		V	*0.07*	0.078	0.023	0.042	0.112
		VI	*0.07*	0.082	0.021	0.049	0.110

Note: * indicates the true parameter lies outside the 89% credible interval of the posterior sample. Under specification VI, σr measures only the residual heterogeneity around the demographic-conditional mean, and the model assigns some of that residual variance to the demographic component.

At the population level, the cost of mis-specification is small. The demographic coefficients in specifications II and IV are recovered accurately even though those specifications incorrectly impose $\sigma_r^2 = 0$. Specification V compensates for the missing demographic structure by inflating σ_r to absorb the demographic variance into residual heterogeneity. Specifications I and III, by assuming homogeneity in r, attribute all the variation in choices to the noise parameter μ, which therefore comes out larger than the true value. None of this is surprising: when you tell a statistical model that one source of variation is shut off, it pushes the variation it sees into the sources you have left open.

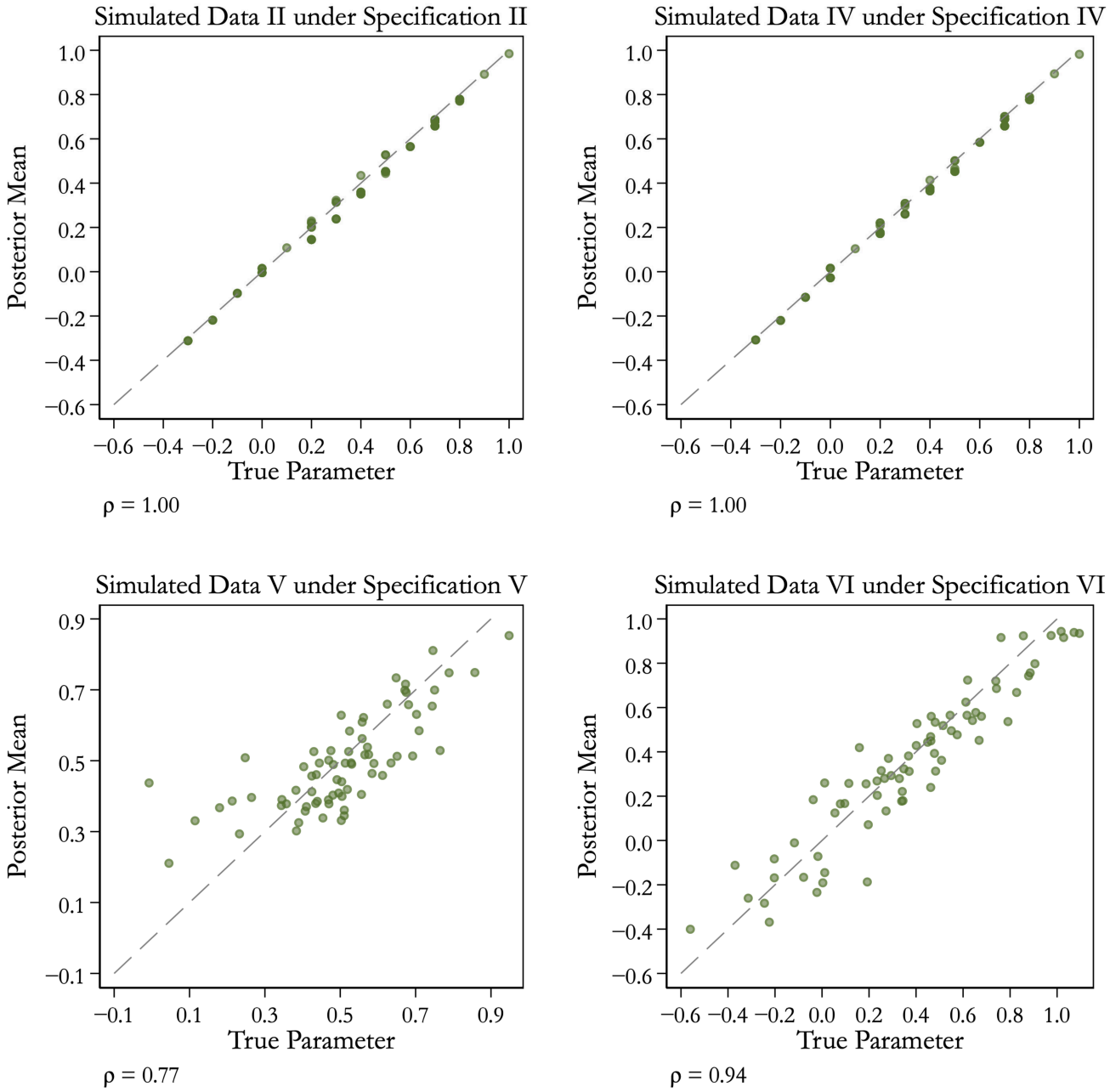


Figure 1. Individual CRRA coefficients in simulated data estimation.

At the individual level, however, the cost of mis-specification is sharper. Figure 2 shows scatter plots of estimated versus true individual r_i under each of the four imperfect specifications applied to dataset VI. The correlation rises monotonically as the specification becomes more correct: 0.87 under II and IV, 0.91 under V, and 0.94 under VI. More tellingly, the credible intervals tell different stories. Under specifications II and IV, which assume $\sigma_r^2 = 0$, the posterior standard deviations of individual r_i are tiny (typically between 0.01 and 0.06), reflecting the false confidence the model places in its (incorrect) assumption of within-cell homogeneity. The result is that, for 53 or 54 of 73 subjects, the true r_i falls outside the 89% credible interval (Table A7). Under specifications V and VI, the posterior standard deviations are larger (between 0.04 and 0.23), and the 89% intervals fail to cover the true value for only about 15 of 73 subjects. The narrow intervals under II and IV are not a feature; they are a symptom of forcing the model to be more certain than the data warrant.

Table 3. Expected utility models using simulated dataset VI under mis-specified specifications.

Group	Parameter	Spec	True	Mean	SD	89% CI Low	89% CI High
r	Constant	***I***	*0.4*	0.432 *	0.019	0.403	0.463
		II	*0.4*	0.419	0.042	0.350	0.485
		III	*0.4*	0.417	0.025	0.377	0.457
		IV	*0.4*	0.395	0.043	0.327	0.460
		V	*0.4*	0.354	0.047	0.282	0.429
		VI	*0.4*	0.359	0.071	0.234	0.463
r	Female	***II***	*0.2*	0.170	0.030	0.124	0.219
		IV	*0.2*	0.179	0.031	0.128	0.227
		VI	*0.2*	0.191	0.056	0.091	0.271
r	Black	***II***	*−0.2*	−0.207	0.027	−0.251	−0.164
		IV	*−0.2*	−0.196	0.027	−0.237	−0.152
		VI	*−0.2*	−0.181	0.058	−0.272	−0.089
r	BusinessMajor	***II***	*−0.5*	−0.509	0.038	−0.568	−0.447
		IV	*−0.5*	−0.503	0.043	−0.569	−0.429
		VI	*−0.5*	−0.488	0.064	−0.595	−0.391
r	HighGPA	***II***	*0.1*	0.128	0.031	0.076	0.175
		IV	*0.1*	0.144	0.029	0.099	0.190
		VI	*0.1*	0.117	0.059	0.024	0.217
r	Insured	***II***	*0.3*	0.269	0.028	0.225	0.313
		IV	*0.3*	0.271	0.028	0.226	0.317
		VI	*0.3*	0.304	0.061	0.207	0.405
r	Std. Dev.	***V***	*0.2*	0.355 *	0.039	0.292	0.414
		VI	*0.2*	0.098 *	0.014	0.076	0.119
μ	Mean	***I***	*0.23*	0.271 *	0.014	0.249	0.295
		II	*0.23*	0.216	0.009	0.200	0.230
		III	*0.23*	0.304 *	0.031	0.255	0.350
		IV	*0.23*	0.232	0.019	0.204	0.261
		V	*0.23*	0.208	0.015	0.184	0.232
		VI	*0.23*	0.210	0.016	0.184	0.234
μ	Std. Dev.	***III***	*0.07*	0.145 *	0.046	0.078	0.213
		IV	*0.07*	0.104	0.025	0.065	0.141
		V	*0.07*	0.083	0.020	0.052	0.111
		VI	*0.07*	0.082	0.021	0.049	0.110

Note: * indicates the true parameter lies outside the 89% credible interval of the posterior sample.

Two specific examples sharpen the point. Subject 54 is moderately risk averse in the true data-generating process, with $r_{54} = 0.329$. Under specification II, the posterior mean is -0.001, classifying her as risk neutral with high (false) confidence; under the correct specification VI, the posterior mean is 0.280. Subject 31 is risk loving in the true data-generating process, with $r_{31} = -0.203$. Under specification V, the posterior mean is 0.150, classifying him as risk averse; under specification VI, the posterior mean is -0.168, correctly identifying him as risk loving. These are not pathological cases. They are the kind of mis-classification that mis-specification produces, and the kind of error that matters when individual risk-preference estimates are inputs to downstream welfare analysis.

The simulation evidence supports a clear conclusion. When the data-generating process is genuinely conditional on covariates, the conditionally exchangeable specification (VI) recovers individual parameters most accurately. When the analyst is unsure whether the data-generating process has demographic structure, the unconditionally exchangeable specification (V) is a robust fallback: it pays a modest cost in correlation (0.91 vs. 0.94 on

dataset VI) but avoids the systematic mis-classification of individual subjects that the more restrictive specifications produce. The specifications that impose $\sigma_r^2 = 0$, II and IV, are the dangerous ones, not because they fail at the population level, but because they convert mis-specification into false confidence at the individual level.

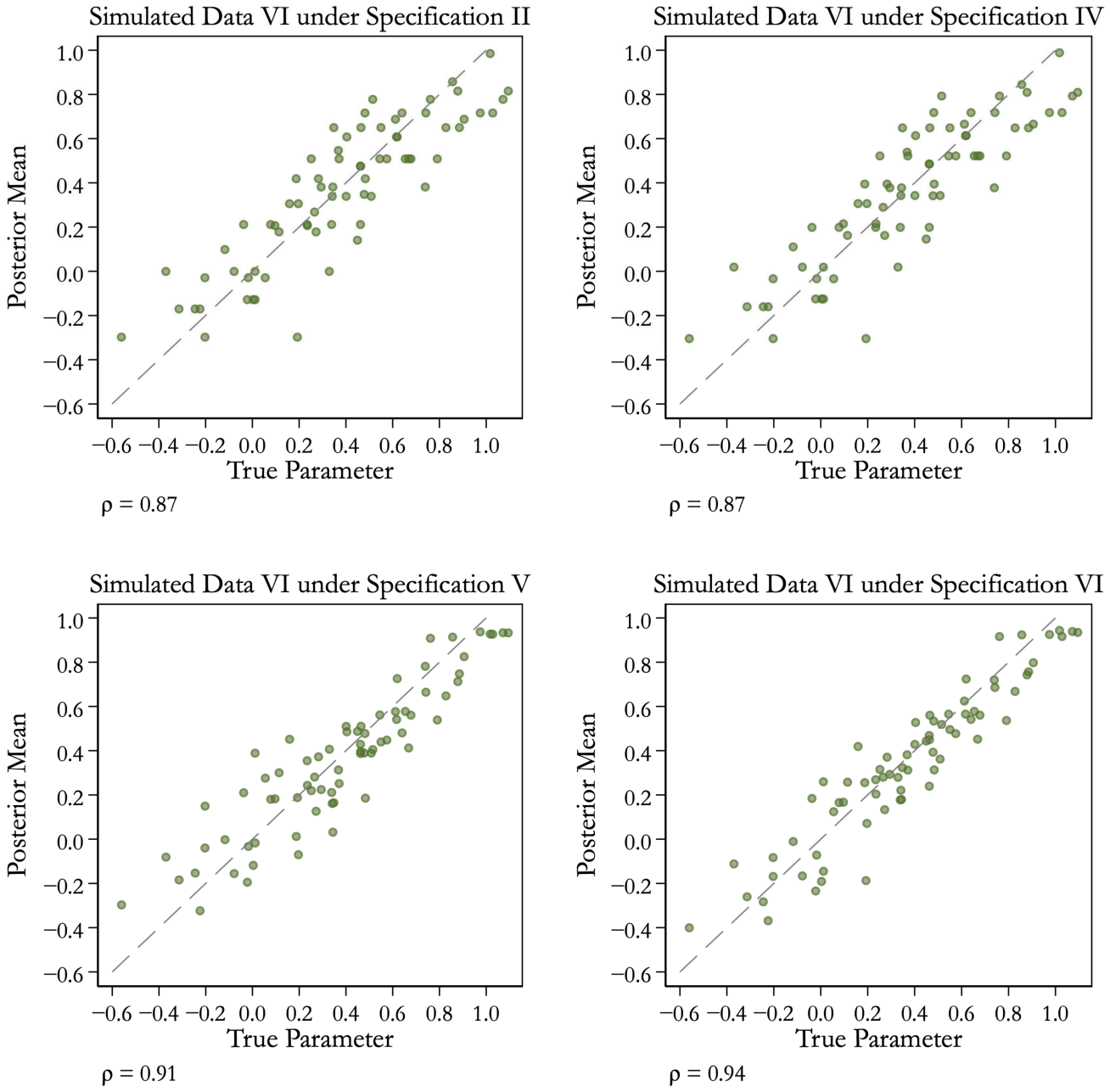


Figure 2. Individual CRRA coefficients using simulated dataset VI assuming incorrect specifications.

3.3. EUT Inferences with Observed Data

We now report EUT estimates from the observed choices of the 73 subjects across the six specifications. All parameters converge with adequate to excellent mixing; the demographic coefficients in specification VI converge more slowly than the others. Population-level estimates are in Table 4.

Table 4. Expected utility estimates with observed choices.

Group	Parameter	Spec	Mean	SD	89% CI Low	89% CI High
r	Constant	***I***	0.575	0.014	0.553	0.597
		II	0.493	0.040	0.430	0.560
		III	0.657	0.015	0.633	0.680
		IV	0.634	0.035	0.581	0.695
		V	0.512	0.035	0.456	0.567
		VI	0.431	0.094	0.272	0.569
r	Female	***II***	0.079	0.028	0.036	0.122
		IV	0.079	0.030	0.036	0.130
		VI	0.070	0.080	−0.055	0.195
r	Black	***II***	0.057	0.030	0.007	0.104
		IV	0.005	0.027	−0.039	0.048
		VI	0.087	0.084	−0.048	0.223
r	BusinessMajor	***II***	−0.077	0.029	−0.122	−0.032
		IV	−0.105	0.031	−0.154	−0.058
		VI	−0.018	0.075	−0.142	0.097
r	HighGPA	***II***	0.075	0.028	0.029	0.117
		IV	0.020	0.030	−0.026	0.067
		VI	0.028	0.080	−0.103	0.149
r	Insured	***II***	−0.046	0.028	−0.087	0.002
		IV	−0.067	0.027	−0.116	−0.027
		VI	−0.068	0.093	−0.207	0.089
r	Std. Dev.	***V***	0.280	0.030	0.232	0.325
		VI	0.164	0.018	0.135	0.190
μ	Mean	***I***	0.232	0.011	0.214	0.250
		II	0.230	0.012	0.212	0.248
		III	0.335	0.052	0.256	0.411
		IV	0.303	0.040	0.240	0.360
		V	0.183	0.011	0.165	0.200
		VI	0.184	0.012	0.165	0.203
μ	Std. Dev.	***III***	0.279	0.093	0.147	0.407
		IV	0.227	0.066	0.128	0.317
		V	0.058	0.015	0.035	0.080
		VI	0.058	0.017	0.032	0.083

Consider first the conditionally exchangeable specification VI, which allows for both demographic and individual heterogeneity. The posterior means of the five demographic coefficients are all close to zero, and the 89% credible intervals all contain zero. Substantial individual heterogeneity remains: the posterior mean of σ_r is 0.164. The histogram of the posterior means of r_i across the 73 subjects is shown in Figure 3. A total of 62 of the 73 subjects have an 89% credible interval for r_i that lies entirely above zero; for the remaining 11 subjects the credible interval contains zero, so the data are consistent with risk neutrality. The Fechner noise parameter μ_i has a posterior median of 0.184 and posterior standard deviation of 0.058.

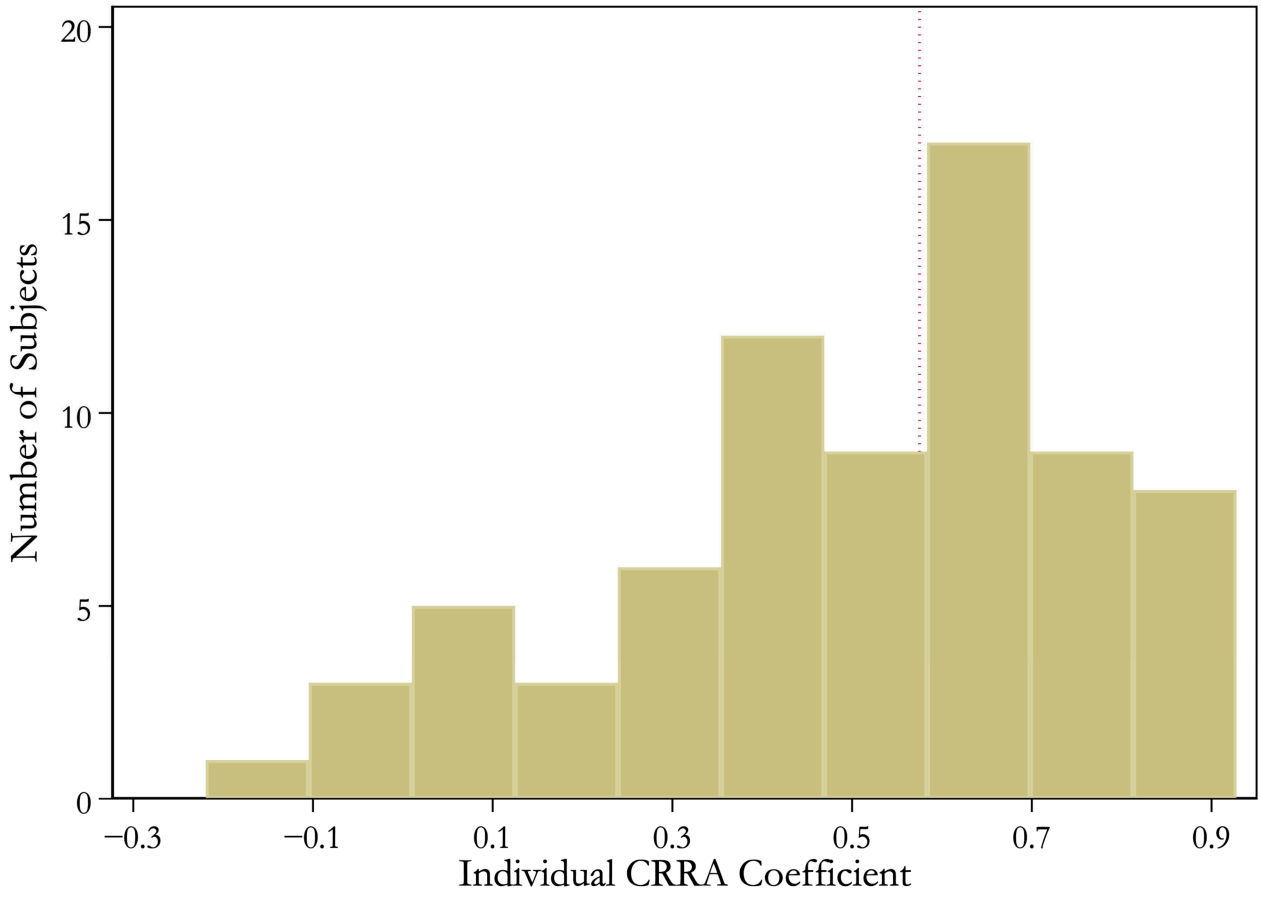


Figure 3. Individual CRRA coefficients r_i from observed choices under EUT specification VI.

Specification V, which drops the demographic structure, produces a very similar picture. The posterior mean of r_i is 0.512 with σ_r = 0.280, and the individual posterior means under V are almost identical to those under VI: the correlation across subjects is essentially 1, as shown in the left panel of Figure 4. The estimates of μ_i are likewise nearly indistinguishable. Specifications II and IV, which impose $\sigma_r^2 = 0$, produce individual estimates that diverge sharply from those of VI, with linear correlation coefficients of around 0.23 (Figure 4, middle and right panels), the same pattern we saw with simulated data. Specifications I and III, which impose homogeneity in r, absorb the variation into μ, which is correspondingly larger.

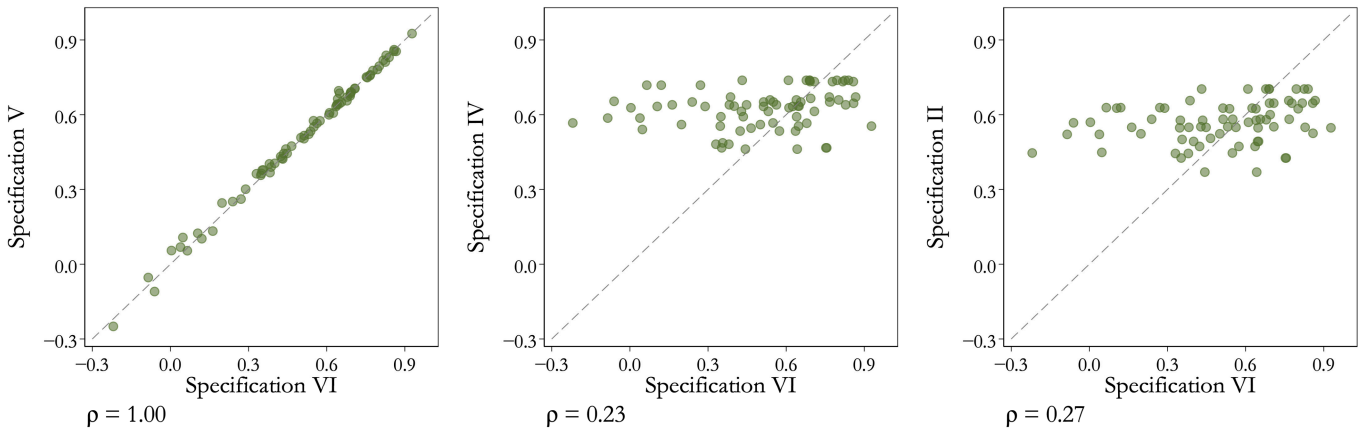


Figure 4. Individual CRRA coefficients r_i from observed choices under different EUT specifications.

We perform pairwise Bayesian comparisons of specifications I through V against VI, using log-Bayes factors. The log-Bayes factor against I, II, III and IV is 221.81, 207.44, 165.10 and 146.45, respectively, indicating that, under the priors we have assumed, the data are much better described by VI than by any of those four. Against specification V, however, the log-Bayes factor is −6.61, going the other way. Posterior model probabilities, computed under the assumption of equal prior weight on each of the six specifications, place 99.9% of the mass on specification V. With the caveats about Bayes factors noted earlier, then, the in-sample evaluation favors V over the alternatives in our sample.

The substantive picture for our population is straightforward. There is no demographic structure to find in this sample: conditional exchangeability buys nothing. The unconditionally exchangeable hierarchical model (V) describes the data as well as the conditionally exchangeable model (VI), at the cost of five fewer estimated coefficients that turn out to be essentially zero anyway. This is precisely the case our discussion of trade-offs anticipates: when there is no demographic structure, the cost of conditioning is small but not zero. For a sample drawn from a more heterogeneous population, a field sample with substantially more demographic variation than our undergraduate subject pool, the comparison could easily go the other way. The Bayesian apparatus is the same; what changes is the data.

4. Rank-Dependent Utility Theory

We extend the analysis to RDU, where the risk premium is decomposed between utility curvature and probability weighting. Under RDU, prizes are ranked from best to worst and the decision weight associated with each prize is the difference of cumulative transformations of the objective probabilities under a probability weighting function. We use the two-parameter Prelec (1998) specification,

$$\omega(p) = exp\left(-\eta(-ln\ p)^{\varphi}\right),$$

with $\eta > 0$ and $\varphi > 0$. The RDU of a lottery is the decision-weighted sum of utilities of prizes, with utility curvature governed by the CRRA parameter r_i as before. The latent index for choice t by subject i is the difference between the RDU of the two lotteries, scaled by the Fechner noise parameter μ_i and a contextual normalization. EUT is nested within RDU when $\eta = \varphi = 1$.

The six specifications carry over directly. For r_i and μ_i we use the same priors as for EUT. For η_i and φ_i, which must be positive, we use log-normal priors. In specification VI the hyper-means are linear in X_i:

$$ln(\eta_i) \sim N\left(\gamma \cdot X_i,\ \sigma^2_{ln\ \eta}\right) \text{and } ln(\varphi_i) \sim N\left(\lambda \cdot X_i,\ \sigma^2_{ln\ \varphi}\right).$$

In specification V the hyper-means are scalar. In specifications II and IV the within-cell variances are set to zero. The four parameters are not independent in the posterior even though we assume independent priors: the RDU model decomposes the risk premium between utility curvature and probability weighting, and the well-known trade-off between these two channels induces posterior correlation among r_i, η_i and φ_i.

We bypass the simulation evidence here, since it is qualitatively identical to the EUT case, and turn to the observed choices. Estimates from all six specifications are in Table 5. The pattern of Bayes factors and posterior model probabilities is essentially the same as under EUT, with the same cautions about what those metrics do and do not tell us. In our sample, specification V again describes the data better than the others under the priors we have assumed. Using V as the baseline, the log-Bayes factor against I, II, III, IV and VI is 300.02, 264.37, 225.48, 204.88 and 12.59, respectively, and the posterior probability

on V is 100% under equal prior weights. Once again, conditional exchangeability on these covariates does not improve in-sample fit in any material way, but it costs little: the estimates from V and VI are close, and the conditional model is correctly identifying that the demographic coefficients are small.

Under the preferred specification V, Figure 5 shows histograms of the posterior means of r_i, η_i, and φ_i across the 73 subjects. There is substantial heterogeneity in all four parameters. A total of 56 subjects have posterior mean φ_i below 1, and 17 have posterior mean φ_i above 1. A comparison of posterior means of r_i under EUT and RDU is shown in Figure 6: the CRRA estimates are slightly higher under EUT, consistent with EUT having to absorb into utility curvature alone what RDU spreads between curvature and probability weighting.

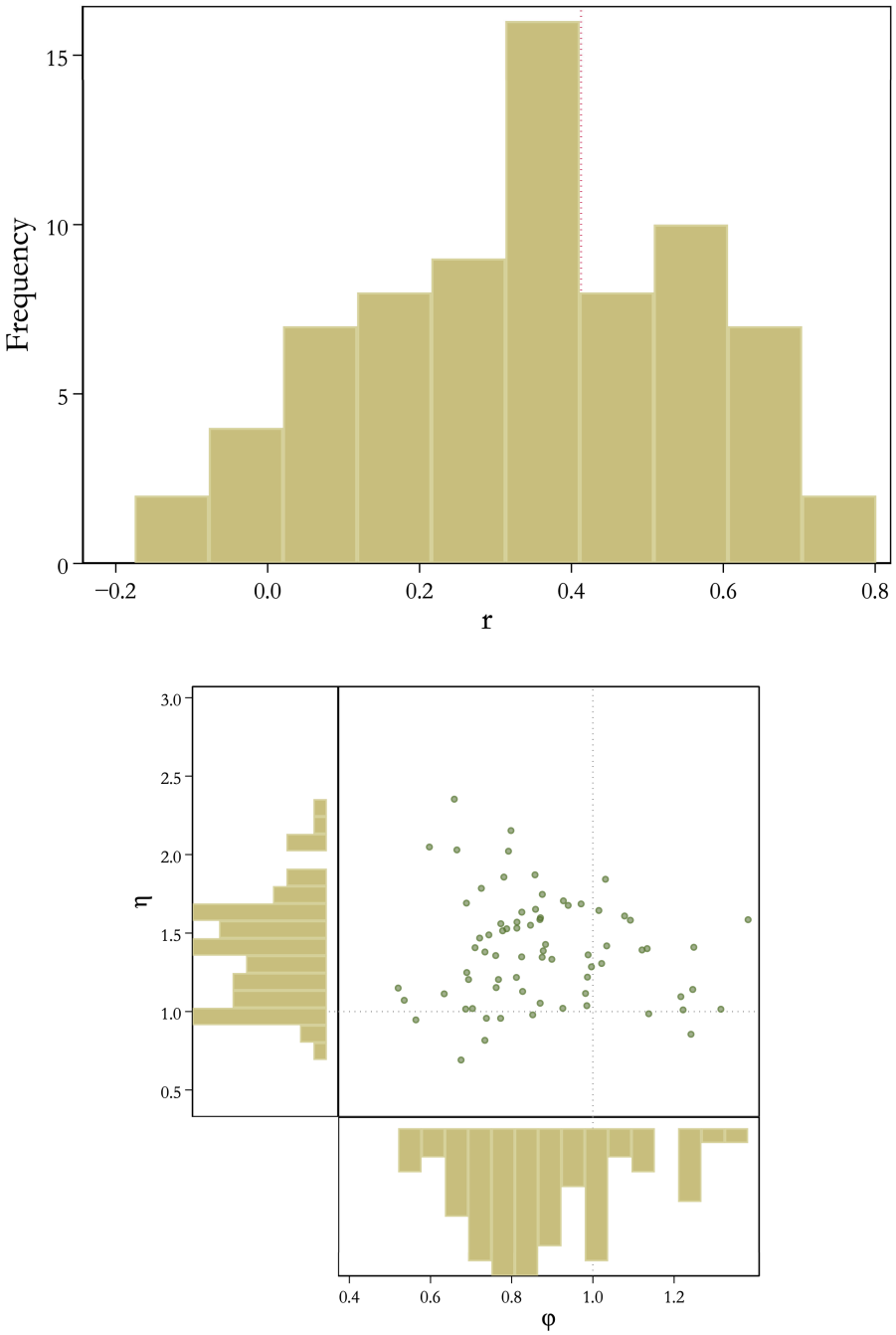


Figure 5. Posterior means of individual estimates assuming Specification V of the RDU Prelec model.

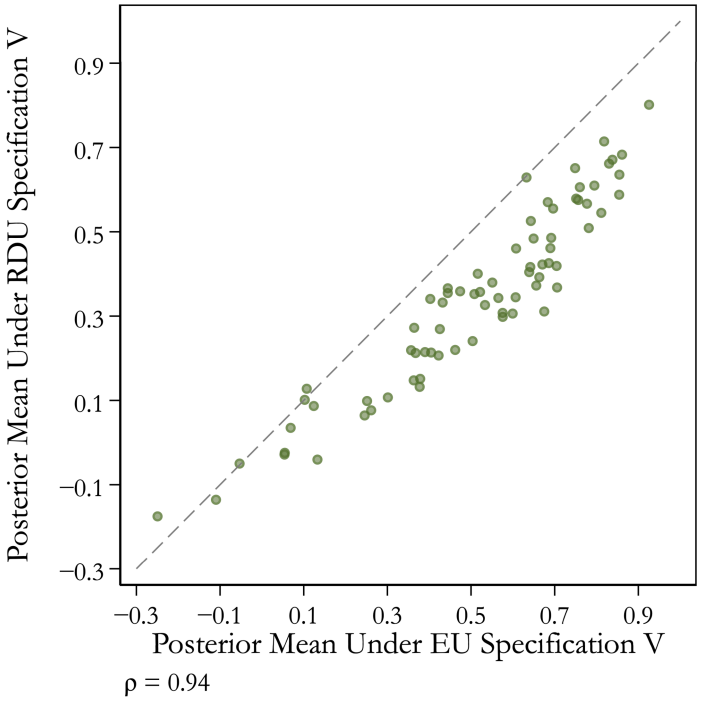


Figure 6. Posterior means of individual CRRA coefficients assuming Specification V of EUT and RDU Prelec model.

The substantive lesson from the RDU estimates is the same as from EUT. Conditional exchangeability and unconditional exchangeability give nearly identical individual-level estimates here, because the demographic covariates carry essentially no signal about risk preferences *in this subject pool*. The in-sample evaluation favors V, and that ranking would plausibly flip in a population where demographic covariates did carry signal. The question is empirical, not conceptual.

Table 5. Rank-Dependent Utility estimates with observed choices.

Group	Parameter	Spec	Mean	SD	89% CI Low	89% CI High
r	Constant	***I***	0.413	0.045	0.338	0.480
		II	0.577	0.104	0.429	0.746
		III	0.503	0.038	0.443	0.564
		IV	0.478	0.103	0.309	0.635
		V	0.343	0.056	0.259	0.435
		VI	0.331	0.109	0.174	0.512
r	Female	***II***	−0.311	0.073	−0.421	−0.188
		IV	−0.113	0.136	−0.321	0.097
		VI	−0.058	0.114	−0.226	0.142
r	Black	***II***	0.056	0.076	−0.062	0.175
		IV	0.082	0.122	−0.118	0.256
		VI	0.085	0.103	−0.074	0.252
r	BusinessMajor	***II***	0.231	0.062	0.134	0.329
		IV	0.103	0.135	−0.092	0.294
		VI	0.003	0.106	−0.148	0.185
r	HighGPA	***II***	−0.182	0.065	−0.289	−0.082
		IV	−0.068	0.075	−0.185	0.059
		VI	−0.079	0.114	−0.266	0.094

Table 5. *Cont.*

Group	Parameter	Spec	Mean	SD	89% CI Low	89% CI High
r	Insured	***II***	−0.040	0.058	−0.132	0.047
		IV	0.011	0.074	−0.103	0.126
		VI	−0.019	0.111	−0.208	0.144
r	Std. Dev.	***V***	0.283	0.045	0.212	0.352
		VI	0.175	0.032	0.127	0.223
η	Constant	***I***	1.270	0.099	1.113	1.425
		II	0.645	0.152	0.381	0.837
		III	1.320	0.109	1.158	1.501
		IV	1.066	0.219	0.710	1.395
		V	1.395	0.126	1.194	1.580
		VI	1.063	0.188	0.757	1.320
η	Female	***II***	0.835	0.137	0.613	1.052
		IV	0.575	0.231	0.189	0.913
		VI	0.404	0.135	0.202	0.641
η	Black	***II***	0.073	0.149	−0.173	0.306
		IV	−0.071	0.213	−0.355	0.354
		VI	−0.013	0.133	−0.208	0.210
η	BusinessMajor	***II***	−0.496	0.133	−0.704	−0.287
		IV	−0.380	0.286	−0.800	0.037
		VI	−0.106	0.134	−0.312	0.112
η	HighGPA	***II***	0.425	0.140	0.194	0.640
		IV	0.097	0.172	−0.166	0.378
		VI	0.166	0.154	−0.101	0.379
η	Insured	***II***	0.091	0.113	−0.092	0.269
		IV	−0.033	0.158	−0.272	0.246
		VI	0.033	0.142	−0.228	0.235
η	Std. Dev.	***V***	0.522	0.145	0.303	0.739
		VI	0.199	0.058	0.117	0.289
φ	Constant	***I***	0.841	0.033	0.790	0.894
		II	0.708	0.076	0.591	0.840
		III	0.817	0.033	0.768	0.870
		IV	0.651	0.065	0.547	0.749
		V	0.876	0.045	0.801	0.944
		VI	0.652	0.073	0.548	0.791
φ	Female	***II***	0.080	0.090	−0.078	0.213
		IV	0.162	0.102	0.004	0.317
		VI	0.167	0.093	0.026	0.319
φ	Black	***II***	0.103	0.079	−0.018	0.229
		IV	0.100	0.086	−0.043	0.233
		VI	0.101	0.111	−0.080	0.259
φ	BusinessMajor	***II***	0.247	0.080	0.124	0.379
		IV	0.168	0.092	0.027	0.319
		VI	0.099	0.109	−0.081	0.266
φ	HighGPA	***II***	−0.129	0.080	−0.244	0.008
		IV	−0.117	0.087	−0.258	0.016
		VI	−0.054	0.097	−0.214	0.106
φ	Insured	***II***	0.164	0.078	0.040	0.287
		IV	0.191	0.078	0.067	0.320
		VI	0.227	0.114	0.043	0.410

Table 5. *Cont.*

Group	Parameter	Spec	Mean	SD	89% CI Low	89% CI High
φ	Std. Dev.	***V***	0.267	0.051	0.186	0.347
		VI	0.112	0.021	0.081	0.144
μ	Mean	***I***	0.234	0.013	0.213	0.254
		II	0.218	0.012	0.199	0.237
		III	0.379	0.068	0.280	0.478
		IV	0.311	0.049	0.236	0.380
		V	0.175	0.011	0.157	0.191
		VI	0.177	0.011	0.160	0.192
μ	Std. Dev.	***III***	0.348	0.127	0.181	0.520
		IV	0.240	0.081	0.126	0.343
		V	0.045	0.018	0.015	0.075
		VI	0.041	0.019	0.008	0.067

5. Limitations and Extensions

Our analysis was deliberately kept simple, so that the trade-off at its center could be seen clearly. That simplicity was bought with several restrictions, and each restriction is a limitation that invites an extension. Here, we discuss those we regard as most important, roughly in order from the specification of the model itself, to the evaluation of alternative specifications, to the data one might confront the model with, and, finally, to the computational demands that many of these extensions share. None of the extensions raises a conceptual difficulty for the framework. The binding constraints are practical, and, as we note at the end, they are increasingly manageable.

5.1. Conditional Exchangeability in the Bayesian Literature

The idea of allowing hyper-parameters to depend on observable covariates is not novel, and we would not want our framing to suggest otherwise. Conditioning group-level parameters on observables has been a workhorse of Bayesian hierarchical modeling at least since Lindley and Smith (1972), the formal study of exchangeability and its partial and conditional variants goes back to de Finetti (1937) and Diaconis and Freedman (1980), and the practice is standard in textbook treatments such as Gelman et al. (2013) and Kruschke (2015). In the Bayesian marketing literature it has been routine for decades: Rossi and Allenby (1993) and Allenby and Rossi (1999) condition household-level parameters on observable characteristics precisely because their applied questions demand individual-level inference.

What we claim as a contribution is therefore not the device itself. It is the characterization of the trade-off the device creates in the specific setting of structural risk preferences. As covariates accumulate, the cells of the implied partition of subjects shrink, and with them the borrowing-of-strength that motivated the hierarchical model in the first place. That trade-off matters quantitatively for the individual-level inferences that are increasingly used as inputs to normative welfare evaluation, and it is the trade-off, rather than the conditioning, that we regard as under-studied.

5.2. Demographic Structure in the Variances

We allowed the demographic covariates to enter the hyper-means of the population distributions, and only the hyper-means. The within-cell standard deviations and the population distribution of the Fechner noise parameter μ were assumed common across demographic groups, a restriction we flagged earlier. We agree that this is a limitation. There is no principled reason why the dispersion of risk preferences or the noisiness of

choice behavior should be identical across demographic groups, and there are settings in which one would positively expect otherwise (e.g., age and education could plausibly affect the consistency of responses to lottery tasks as much as their central tendency). If one group is genuinely more dispersed than another, a model that imposes a common within-cell variance will shrink the members of the more dispersed group too aggressively toward their cell mean, a milder cousin of the false confidence we documented for specifications that impose zero within-cell variance.

The extension is conceptually immediate: let the log of each within-cell standard deviation be linear in the same covariates that enter the hyper-means, and let the data decide. Our reason for not doing so here was tractability, not principle. Conditioning the variances as well as the means doubles the number of hyper-parameters per covariate, and raises the identification demands on a sample of the size used here. A useful first step, short of full estimation, would be simulations in which the group-level variances are deliberately uneven, to establish how much damage the common-variance restriction does to individual-level recovery when it is false.

5.3. Interactions, Non-Linearities and Irrelevant Covariates

Our covariates entered the hyper-means linearly, and singly. Richer covariate structures are natural to consider, such as interactions between covariates, and non-linear terms such as quadratics (which are just interactions of a covariate with itself). The trade-off we characterize applies to these with extra force, because interactions multiply cells faster than main effects do. Five binary covariates define up to 32 cells; their pairwise interactions define many more, so the pooling erodes correspondingly faster.

A complementary exercise, which our simulation design makes straightforward, is to include covariates that are known by construction to carry no signal at all, so that they contribute nothing but noise. Our comparison of specifications V and VI with observed data suggests that the price of a handful of such covariates is small, since the posterior simply concludes that they do not matter. Whether that reassurance survives many noise covariates, or noise covariates interacted with informative ones, is exactly the kind of question that simulated data with known truth are suited to answer, and it is a direct extension of the design used here.

5.4. Richer Structural Models of Risk Preferences

Our analysis considered EUT and RDU specifications. The extension to richer structural models, most obviously Cumulative Prospect Theory (CPT), is important. A full CPT specification asks a great deal of the data of each subject: distinct utility curvatures over gains and over losses, a loss aversion parameter, and distinct probability weighting over gains and over losses, all identified from choices that must span gain, loss and mixed frames. That is exactly the setting in which the informative priors of a BHM do their greatest work, a point made explicitly by Harrison and Swarthout (2023, p. 160) in proposing BHMs as the constructive path for the econometrics of CPT, and demonstrated by Bland et al. (2026), who evaluate loss aversion at the level of the individual, rather than for the average or median individual, across multiple datasets. The hierarchical estimation of CPT models by Nilsson et al. (2011) reflects the same motivation. By the logic developed here, CPT models should also benefit from demographic conditioning of the hyper-parameters, up to a point, and that point is likely to arrive sooner than it does for EUT. For CPT, each additional covariate now enters the hyper-means of many individual-level parameters at once, so the proliferation of hyper-parameters, and the attendant erosion of pooling, is faster. The trade-off we characterize does not vanish in richer models; it binds more tightly.

5.5. Alternatives to Bayes Factors

Our evaluations of alternative specifications relied on Bayes factors and posterior model probabilities, with the cautions set out earlier. One might ask whether other comparison criteria, such as in-sample information criteria, or predictive measures such as the Widely Applicable Information Criterion and leave-one-out cross-validation reviewed by Vehtari et al. (2017), would tell a different story.

We flag the possibility deliberately, and just as deliberately we counsel caution. Many information criteria amount to *ad hoc* punishments for simply adding covariates, even when dressed up as having some Bayesian rationale. The BIC of Schwarz (1978) is a familiar example, and it penalizes exactly the covariate additions whose empirical value is the question at issue here, on asymptotic grounds that have nothing to do with the priors actually used. When the central question is whether added covariates earn their keep, a criterion that mechanically taxes them assumes the answer. Our general position follows the discussion of model comparison in Harrison and Rutström (2009): comparative measures of fit should be chosen to match the inferential purpose at hand, and not generated mechanically because everyone else does so. Where the purpose is out-of-sample prediction, predictive criteria are the appropriate tool, and evaluating them within our simulation design, where the truth is known, would be a worthwhile extension. Where the purpose is the normative use of individual estimates, the constrained specification searches discussed earlier remain the appropriate discipline.

5.6. The Assumed Stability of Covariate Effects

Everything we do assumes that the mapping from covariates to preferences, the hyper-coefficients β, is a stable feature of the population over the horizon of the inference. That assumption becomes substantive the moment the estimates travel to the same population at a later date, or to the wider survey population considered below. It deserves a careful statement, because the literature on the stability of risk preferences has often muddled two distinct notions. Unconditional stability asks whether elicited preferences are the same when measured at two points in time, with no allowance for anything else having changed in between. Conditional stability asks whether preferences are stable once changes in the relevant states of nature are conditioned on. Andersen et al. (2008) flag and illustrate the distinction with longitudinal field data: elicited risk attitudes varied over a 17-month span, but with no general drift, and the variation was associated with changes in the personal financial circumstances of subjects. Instability in the demanding unconditional sense, which is what many claims of "unstable preferences" amount to, is entirely consistent with stability in the weaker conditional sense.

It is the conditional sense that our framework assumes, since the covariates are inside the model. But an assumption is not a finding. Where longitudinal data with demographic covariates exist, one can estimate directly whether the hyper-coefficients are conditionally stable over time, and we would encourage exactly such studies before covariate-conditioned estimates are transported across time or across populations.

5.7. Conditioning in the Field: Combining Survey and Experimental Data

From a methodological standpoint, the case for routinely allowing some conditioning on observables is reasonable when there is *a priori* reason to think the chosen covariates could matter. From a practical standpoint, Bayes factors and posterior model probabilities, used as evaluations rather than verdicts, with all the cautions discussed earlier, provide a useful summary of whether the conditioning has improved the in-sample description of the data. In our specific sample of undergraduate subjects, that evaluation places overwhelming weight on the unconditionally exchangeable specification, because the

demographic covariates here carry essentially no signal about risk preferences. In a more heterogeneous field sample, the evaluation could easily run the other way. The apparatus is the same; the conclusion is empirical.

One extension is natural, and perhaps the most consequential for applied work. The framework developed here makes it possible to combine cheap, large-sample survey data on a field population with expensive, incentivized experimental data on a small sub-sample drawn from that population, in a way that delivers risk-preference estimates with the demographic reach of the survey and the inferential discipline of the experiment. The design we have in mind is straightforward in principle. A large representative survey of the target field population collects the demographic covariates $\mathbf{X}$ and a battery of responses to risk-preference questions that have the same structural form as the experimental tasks used here. Typically, these might be choices between paired lotteries or switch points in a multiple price list, posed hypothetically but otherwise identical in their incentive structure to the incentivized versions. A random sub-sample of the survey respondents is then invited to a paid laboratory or field experiment in which they make the same kinds of choices over real monetary stakes, with their demographics again recorded. The experimental sub-sample yields incentivized estimates of individual risk preferences for the subjects who participated; the survey, taken alone, yields hypothetical responses for the full sample. The Bayesian apparatus is what links them.

Before introducing the specific ways of linking the two, we should note that the insistence on structural similarity between the field survey and the laboratory experiment is doing a great deal of work here, and that it is worth being explicit about what it rules out. Survey instruments that ask respondents to rate their willingness to take risks on a Likert scale, the form popularized by Dohmen et al. (2011) and reproduced widely in subsequent applied work, are not what we have in mind, and would not serve the purpose envisaged here. As Harrison (2025) documents, such Likert-style instruments are subject to at least three distinct problems that compound rather than offset one another. The first is hypothetical bias: responses to risk questions without real stakes need not, and demonstrably do not, line up with responses to incentivized choices over the same lotteries, with the inferred risk-preference parameters differing substantially even when the two instruments are administered to the same subjects. The second is the loss of theoretical content: a position on a 0–10 Likert scale is not the latent construct that the thought experiment defining risk preferences refers to, and mapping such a scale to a CRRA coefficient (or to any other parameter in a structural utility model) requires a non-linear transformation that the linear regressions used in the relevant literature simply ignore. The third is the temptation, once an analyst has the data in hand, to combine several such measures by factor analysis or principal components and treat the resulting score as a measure of risk preferences; this compounds the first two problems, imposing a statistical structure unrelated to any theoretical thought experiment about risk and collapsing what may be genuinely heterogeneous information into a single index whose connection to a structural parameter is, at best, decorative. For our design to do what we want it to do, then, the survey instrument must ask the same kind of question as the experimental task, a choice between two lotteries or a switch point in a multiple price list, without the financial incentive but with the structural identification preserved, so that the same thought experiment is in force on both sides of the comparison.

The natural way to do that linking, in the spirit of the conditionally exchangeable specification, is to estimate the hierarchical model on the experimental sub-sample with the demographic covariates entering the hyper-parameters, and then use the posterior of the hyper-coefficients β, γ and λ, together with the demographic profile of the wider survey population, to project the distribution of risk preferences out to the non-experimental survey

respondents. Two things would matter regarding whether that projection is informative. The first is whether the demographic covariates in fact carry signals about risk preferences in this population, which is exactly the question we framed as the central empirical issue. In our undergraduate sample they did not; in a field population with more demographic variation they plausibly will. The second is whether the hypothetical survey responses are themselves informative about incentivized risk preferences for the same individuals, which can be tested directly within the experimental sub-sample by comparing the two responses and using the result to discipline, or, where appropriate, to discard, the survey signal from the wider sample.[7]

If the comparison reveals systematic discrepancy rather than statistical independence between hypothetical and incentivized responses, the appropriate response is not to throw the survey out but to model the discrepancy. The idea, developed by Blackburn et al. (1994) in the context of contingent-valuation surveys, is to treat the hypothetical bias as a function of the observable characteristics of the respondent, to estimate that function on data where both kinds of responses are available, and to use it to calibrate the hypothetical responses elsewhere. In their original application, the bias function was estimated from one good and applied to a different good, with encouraging out-of-sample performance from both parametric and semiparametric specifications. The same logic applies here, and in a setting that is in one respect more favorable: our experimental sub-sample provides hypothetical and incentivized responses for the same individuals over the same lotteries, with the same demographic vector X, so the bias function can be estimated within-sample rather than transferred across goods. The bias-corrected hypothetical responses from the wider survey then enter the hierarchical model as demographically anchored input alongside, or in place of, the raw survey responses. Whether such calibration buys much, beyond what the covariate-in-hyper-parameter approach already delivers, depends on how much independent signal the hypothetical responses carry once the demographics have been conditioned on. That is a question we can ask of the data rather than answer in advance.

A second, simpler point is worth making. Almost all of the methodological work on hierarchical Bayesian estimation of risk preferences, including the analysis above, has used data from laboratory samples drawn from undergraduate populations. The demographic range of such samples is, by construction, narrow, and our finding that the conditioning buys little in this sample is in part an artefact of that narrowness. Field populations of the kind used in Harrison et al. (2007), where the same risk-elicitation procedure was carried out across Denmark with a nationally representative sample spanning ages 19 to 75, education levels from primary school to long-cycle higher education, and substantial geographic variation, are a much more demanding test of the framework. There, the *marginal* effects of education and age on risk preferences are large and statistically well identified, and the *total* effects larger still. Applying the framework developed here to such data, rather than to laboratory samples, is the obvious next step, and the one where the comparison between unconditional and conditional exchangeability is most likely to deliver substantively different individual-level estimates. The apparatus is the same; the demographic structure is what changes.

5.8. Computational Implications

Most of these extensions share a practical obstacle, the speed of computation. Our estimations were carried out in *Stata*, and for the specifications considered here that was adequate. But covariates in the variances, interacted covariates, and CPT-scale parameter vectors all increase the dimension of the posterior and complicate its geometry, and MCMC run-times that are tolerable for a two-parameter EUT model can become prohibitive. That is not a reason to avoid the extensions; it is a reason to change computational tools.

We have ported BHM code of the kind used here to *Stan* (Version 2.37), the probabilistic programming language of Carpenter et al. (2017), and, more consequentially, to *NumPyro*, the probabilistic programming module of Phan et al. (2019), which runs under *Python* (Version 3.14.6) on the JAX framework. *NumPyro* uses the same Hamiltonian Monte Carlo sampler as *Stan*, the No-U-Turn Sampler of Hoffman and Gelman (2014), but compiles the log-posterior and its gradient to run on graphics processing units, and for likelihoods with the embarrassingly parallel structure of binary choice data, the speed improvements are dramatic. Nor does any of this require abandoning a *Stata*-centered workflow: the *Python* code can be invoked from within *Stata* using the Python prefix command, so that data management and post-estimation analysis remain where they were. The porting effort is modest, and validation is direct, since the same model can be estimated on the same data in both environments and the posteriors compared. The computational path for the extensions sketched above therefore exists, and it is feasible.

6. Conclusions

Hierarchical Bayesian estimation of individual risk preferences is a powerful tool when the analyst needs reliable, individual-level estimates from data that, taken one subject at a time, are too sparse to identify each subject's parameters precisely. The hierarchy delivers its gains by pooling: each subject's posterior borrows strength from the rest of the sample through a shared population distribution. The standard implementation assumes that subjects are unconditionally exchangeable, which is what gives the pooling its bite.

Conditional exchangeability is an alternative that allows the population distribution to depend on observable covariates. Its appeal is that it makes use of information the analyst has anyway. Its risk is that it can carve the population into cells so small that the pooling, and with it the principal benefit of the hierarchical model, disappears. Our contribution is to characterize that trade-off concretely, with simulated data where the truth is known and with observed data from a laboratory experiment.

Three findings stand out. First, when the data-generating process actually has demographic structure, a conditionally exchangeable model with a parsimonious set of covariates recovers individual risk preferences better than its unconditionally exchangeable counterpart. Second, when the data-generating process has no demographic structure, the cost of including a few covariates anyway is small: the posterior simply concludes that the covariates do not move the estimates, and the individual-level inferences are essentially the same as under unconditional exchangeability. Third, specifications that impose zero within-cell variance, conditional exchangeability without any residual individual heterogeneity, are dangerous. They produce tight credible intervals that are systematically wrong about the parameters of many subjects. The lesson is that within-cell variance is not a luxury; it is what protects the model from being too confident in the wrong place.

(1) The first six columns correspond to β from the distribution of r_i, and the seventh column corresponds to σ_r^2. For ease of interpretation, we list the mean and standard deviation of μ_i rather than $m_{ln\ \mu}$ and $\sigma^2_{ln\ \mu}$ as the last two columns here and in the presentation of estimation results.

(2) Compared to specification VI, specifications I–IV assume the standard deviation of the normal distribution of r_i is 0, and specifications I, III and V assume the demographic coefficients of the mean are 0. In addition, specifications I and II assume the standard deviation of the log-normal distribution of μ_i is 0.

(3) For ease of reading, we display the simulation parameters in italics to differentiate them from estimated values in all the ensuing tables.

Author Contributions: Conceptualization, X.S.G. and G.W.H.; methodology, X.S.G. and G.W.H.; software, X.S.G. and G.W.H.; validation, X.S.G. and G.W.H.; formal analysis, X.S.G. and G.W.H.; investigation, X.S.G. and G.W.H.; resources, X.S.G. and G.W.H.; data curation, G.W.H.; writing—original draft preparation, X.S.G. and G.W.H.; writing—review and editing, X.S.G. and G.W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original data presented in the study are openly available in OSF at https://osf.io/x6q4a/files/osfstorage/6a6a50d6ea6d36878992cbd5, accessed on 27 July 2026.

Acknowledgments: We are grateful for constructive comments from two referees.

Conflicts of Interest: The authors declare no conflict of interest.

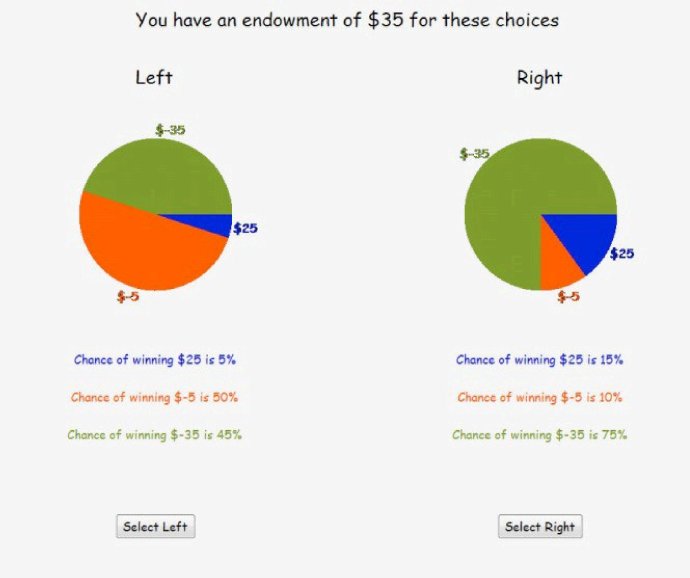
Appendix A. Experiment Instructions and Lottery Parameters (ONLINE)

Appendix A.1. Instructions in the Gain Frame

Choices Over Risky Prospects

This is a task where you will choose between prospects with varying prizes and chances of winning each prize. You will be presented with a series of pairs of prospects where you will choose one of them. For each pair of prospects, you should choose the prospect you prefer. You will actually get the chance to play one of these prospects for earnings, and you will be paid according to the outcome of that prospect, so you should think carefully about which prospect you prefer on each decision screen.

Here is an example of what the computer display of such a pair of prospects will look like.



The outcome of the prospects will be determined by the draw of a random number between 1 and 100. Each number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the number yourself using two 10-sided dice.

You might be told your cash endowment for each decision at the top of the screen. In this example it is $35, so any earnings would be added to or subtracted from this endowment. The endowment may change from choice to choice, so be sure to pay attention to it. The endowment you are shown only applies for that choice.

In this example, the left prospect pays twenty-five dollars ($25) if the number drawn is between 1 and 5, pays negative five dollars ($−5) if the number is between 6 and 55, and pays negative thirty-five dollars ($−35) if the number is between 56 and 100. The blue color

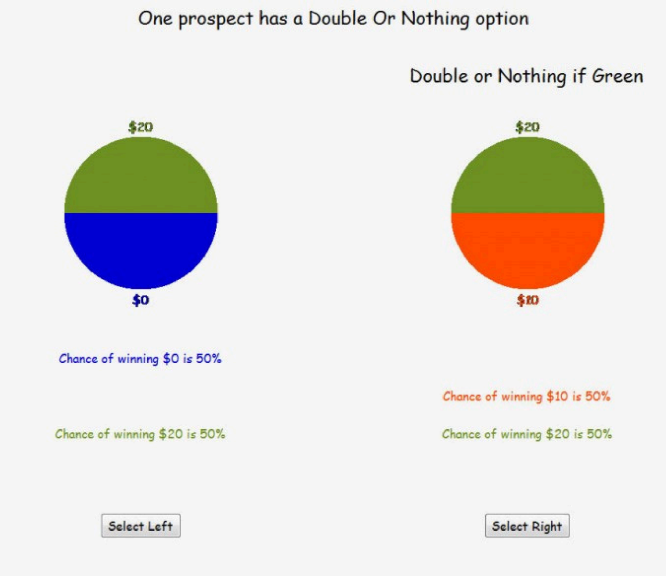
in the pie chart corresponds to 5% of the area and illustrates the chances that the number drawn will be between 1 and 5 and your prize will be $25. The orange area in the pie chart corresponds to 50% of the area and illustrates the chances that the number drawn will be between 6 and 55 and your prize will be $−5. The green area in the pie chart corresponds to 45% of the area and illustrates the chances that the number drawn will be between 56 and 100. When you select the decision screen to be played out, the computer will confirm the die rolls that correspond to the different prizes.

Now, look at the pie on the right. It pays twenty-five dollars ($25) if the number drawn is between 1 and 15, negative five dollars ($−5) if the number is between 16 and 25, and negative thirty-five dollars ($−35) if the number is between 26 and 100. As with the prospect on the left, the pie slices represent the fraction of the possible numbers which yield each payoff. For example, the size of the $25 pie slice is 15% of the total pie.

Even though the screen says that you might win a negative amount, this is actually a loss to be deducted from your endowment. So if you win $−5, your earnings would be $30 = $35 − $5.

Each pair of prospects is shown on a separate screen on the computer. On each screen, you should indicate which prospect you prefer to play by clicking on one of the buttons beneath the prospects.

Some decision screens could also have a pair of prospects in which one of the prospects will give you the chance for "Double or Nothing". For instance, the right prospect in this screen image pays "Double or Nothing" if the green area is selected, which happens if the number drawn is between 51 and 100. The right pie chart indicates that if the number is between 1 and 50 you get $10. However, if the number is between 51 and 100 we will flip a coin with you to determine if you get either double the amount or $0. In this example, if it comes up Heads you get $40, otherwise you get nothing. The prizes listed underneath each pie refer to the amounts before any "Double or Nothing" coin toss.



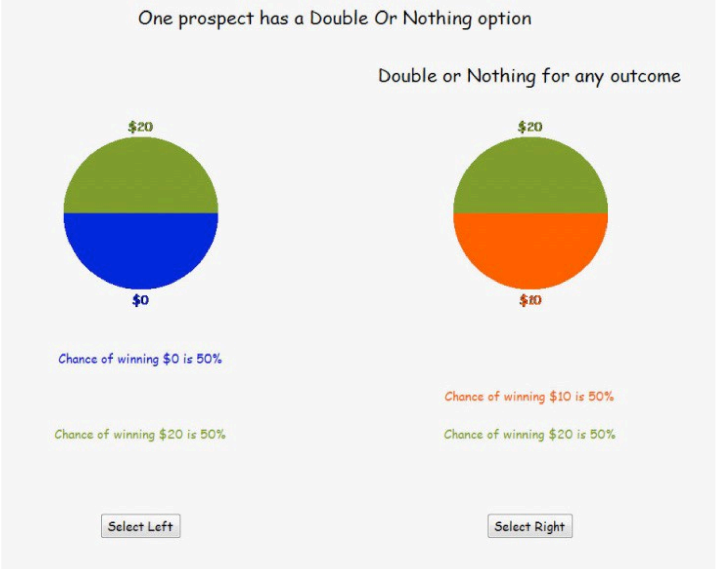
After you have worked through all of the pairs of prospects, please wait quietly until further instructions. When it is time to play this task out for earnings, you will then roll two 10-sided dice until a number comes up to determine which pair of prospects will be played out. If there are 40 pairs we will roll the dice until a number between 1 and 40 comes up, if there are 80 pairs we will roll until a number between 1 and 80 comes up, and so on. Since there is a chance that any of your choices could be played out for real, you should approach each pair of prospects as if it is the one that you will play out. Finally, you will roll the two

ten-sided dice to determine the outcome of the prospect you chose, and if necessary we will then toss a coin to determine if you get "Double or Nothing".

Here is an example: suppose your first roll was 81. We would then pull up the 81st decision that you made and look at which prospect you chose—either the left one or the right one. Let us say that the 81st lottery was the same as the last example, and you chose the left prospect. If the random number from your second roll was 37, you would win $0; if it was 93, you would get $20.

If you picked the prospect on the right and drew the number 37, you would get $10; if it was 93, we would have to toss a coin to determine if you get "Double or Nothing". If the coin comes up Heads then you would get $40. However, if it comes up Tails you would get nothing from your chosen prospect.

It is also possible that you will be given a prospect in which there is a "Double or Nothing" option regardless of the outcome of the random number. This screen image illustrates this possibility.



In summary, your payoff is determined by five things:

- By your endowment, if there is one, shown at the top of the screen;
- By which prospect you selected, the left or the right, for each of these pairs;
- By which prospect pair is chosen to be played out in the series of pairs using the two
- 10-sided dice;
- By the outcome of that prospect when you roll the two 10-sided dice;
- By the outcome of a coin toss if the chosen prospect outcome is of the "Double or Nothing" type.

Which prospects you prefer is a matter of personal choice. The people next to you may be presented with different prospects, and may have different preferences, so their responses should not matter to you or influence your decisions. Please work silently, and make your choices by thinking carefully about each prospect.

All payoffs are in cash, and are in addition to the $5 show-up fee that you receive just for being here, as well as any other earnings in other tasks from the session today.

Appendix A.2. Lottery Parameters

Table A1. Parameters for the lotteries based on Loomes and Sugden (1998).

Lottery ID	Left Lottery						Right Lottery					
	Prize 1	Prob 1	Prize 2	Prob 2	Prize 3	Prob 3	Prize 1	Prob 1	Prize 2	Prob 2	Prize 3	Prob 3
ls2_lr	$10	0.3	$30	0	$50	0.7	$10	0.15	$30	0.25	$50	0.6
ls6_lr	$10	0.6	$30	0	$50	0.4	$10	0	$30	1	$50	0
ls7_lr	$10	0.6	$30	0	$50	0.4	$10	0.15	$30	0.75	$50	0.1
ls10_lr	$10	0.5	$30	0	$50	0.5	$10	0.1	$30	0.8	$50	0.1
ls13_rl	$10	0.5	$30	0.4	$50	0.1	$10	0.7	$30	0	$50	0.3
ls15_rl	$10	0.4	$30	0.6	$50	0	$10	0.5	$30	0.4	$50	0.1
ls17_lr	$10	0.1	$30	0	$50	0.9	$10	0	$30	0.25	$50	0.75
ls18_rl	$10	0.1	$30	0.75	$50	0.15	$10	0.4	$30	0	$50	0.6
ls21_lr	$10	0.7	$30	0	$50	0.3	$10	0.6	$30	0.25	$50	0.15
ls26_rl	$10	0.2	$30	0.6	$50	0.2	$10	0.4	$30	0	$50	0.6
ls29_rl	$10	0.5	$30	0.3	$50	0.2	$10	0.6	$30	0	$50	0.4
ls32_rl	$10	0.7	$30	0.3	$50	0	$10	0.8	$30	0	$50	0.2
ls34_rl	$10	0.1	$30	0.6	$50	0.3	$10	0.25	$30	0	$50	0.75
ls35_rl	$10	0	$30	1	$50	0	$10	0.25	$30	0	$50	0.75
ls39_rl	$10	0.5	$30	0.2	$50	0.3	$10	0.55	$30	0	$50	0.45
ls1i_lr	$10	0.12	$30	0.05	$50	0.83	$10	0.03	$30	0.2	$50	0.77
ls3i_lr	$10	0.27	$30	0.05	$50	0.68	$10	0.03	$30	0.45	$50	0.52
ls7i_lr	$10	0.54	$30	0.1	$50	0.36	$10	0.18	$30	0.7	$50	0.12
ls9i_lr	$10	0.08	$30	0.04	$50	0.88	$10	0.05	$30	0.1	$50	0.85
ls13i_lr	$10	0.65	$30	0.1	$50	0.25	$10	0.55	$30	0.3	$50	0.15
ls16i_lr	$10	0.88	$30	0.04	$50	0.08	$10	0.83	$30	0.14	$50	0.03
ls17i_rl	$10	0.04	$30	0.15	$50	0.81	$10	0.08	$30	0.05	$50	0.87
ls18i_rl	$10	0.14	$30	0.65	$50	0.21	$10	0.38	$30	0.05	$50	0.57
ls22i_lr	$10	0.66	$30	0.1	$50	0.24	$10	0.54	$30	0.4	$50	0.06
ls28i_rl	$10	0.12	$30	0.84	$50	0.04	$10	0.18	$30	0.66	$50	0.16
ls30i_rl	$10	0.45	$30	0.45	$50	0.1	$10	0.55	$30	0.15	$50	0.3
ls31i_lr	$10	0.48	$30	0.36	$50	0.16	$10	0.42	$30	0.54	$50	0.04
ls35i_lr	$10	0.2	$30	0.2	$50	0.6	$10	0.1	$30	0.6	$50	0.3
ls36i_rl	$10	0.02	$30	0.92	$50	0.06	$10	0.08	$30	0.68	$50	0.24
ls37i_lr	$10	0.48	$30	0.28	$50	0.24	$10	0.44	$30	0.44	$50	0.12

Table A2. Parameters for double or nothing lotteries.

Lottery ID	Left Lottery						Right Lottery					
	Prize 1	Prob 1	Prize 2	Prob 2	Prize 3	Prob 3	Prize 1	Prob 1	Prize 2	Prob 2	Prize 3	Prob 3
rdon1	$0	0.5	$10	0.5	$20	0	$0	0.5	$10	0.5	$20	0
rdon2	$0	0	$10	1	$20	0	$0	0.5	$10	0.5	$20	0
rdon3	$0	0	$10	1	$35	0	$0	0	$5	0.5	$18	0.5
rdon4	$0	0.25	$10	0.75	$70	0	$0	0	$35	1	$70	0
rdon5	$0	0	$10	1	$70	0	$0	0	$35	1	$70	0
rdon6	$0	0	$20	1	$35	0	$0	0	$10	0.5	$35	0.5
rdon7	$0	0	$20	0.5	$70	0.5	$0	0	$35	0.5	$70	0.5
rdon8	$0	0	$35	1	$70	0	$0	0	$35	0.5	$70	0.5
rdon9	$0	0	$20	0.5	$35	0.5	$0	0.5	$20	0	$70	0.5
rdon10	$0	0	$35	0.75	$70	0.25	$0	0	$35	1	$70	0
rdon11	$0	0	$20	1	$70	0	$0	0	$20	0.5	$35	0.5
rdon12	$0	0	$35	0.75	$70	0.25	$0	0	$35	0.5	$70	0.5
rdon13	$0	0.25	$10	0.75	$35	0	$0	0.5	$18	0.5	$35	0
rdon14	$0	0	$20	0.75	$35	0.25	$0	0	$18	0.5	$35	0.5
rdon15	$0	0	$20	0.75	$70	0.25	$0	0	$35	0.5	$70	0.5

Table A3. Parameters for the actuarially-equivalent lotteries.

Lottery ID	Left Lottery						Right Lottery					
	Prize 1	Prob 1	Prize 2	Prob 2	Prize 3	Prob 3	Prize 1	Prob 1	Prize 2	Prob 2	Prize 3	Prob 3
rae1	$0	0.5	$10	0.5	$20	0	$0	0.75	$10	0	$20	0.25
rae2	$0	0	$10	1	$20	0	$0	0.75	$10	0	$20	0.25
rae3	$0	0	$10	1	$35	0	$0	0.5	$10	0.25	$35	0.25
rae4	$0	0.25	$10	0.75	$70	0	$0	0.5	$10	0	$70	0.5
rae5	$0	0	$10	1	$70	0	$0	0.5	$10	0	$70	0.5
rae6	$0	0	$20	1	$35	0	$0	0.25	$20	0.25	$35	0.5
rae7	$0	0	$20	0.5	$70	0.5	$0	0.25	$20	0	$70	0.75
rae8	$0	0	$35	1	$70	0	$0	0.25	$35	0	$70	0.75
rae9	$0	0.25	$20	0.5	$70	0.25	$0	0.5	$20	0	$70	0.5
rae10	$0	0	$35	0.75	$70	0.25	$0	0.5	$35	0	$70	0.5
rae11	$0	0	$20	1	$70	0	$0	0.25	$20	0.5	$70	0.25
rae12	$0	0	$35	0.75	$70	0.25	$0	0.25	$35	0	$70	0.75
rae13	$0	0.25	$10	0.75	$35	0	$0	0.75	$10	0	$35	0.25
rae14	$0	0	$20	0.75	$35	0.25	$0	0.25	$20	0	$35	0.75
rae15	$0	0	$20	0.75	$70	0.25	$0	0.25	$20	0	$70	0.75

Table A4. Text for double or nothing lotteries.

Lottery ID	Double or Nothing Text
rdon1	Double or Nothing if outcome 2 in right lottery
rdon2	Double or Nothing if outcome 2 in right lottery
rdon3	Double or Nothing for any outcome in right lottery
rdon4	Double or Nothing for any outcome in right lottery
rdon5	Double or Nothing for any outcome in right lottery
rdon6	Double or Nothing if outcome 2 in right lottery
rdon7	Double or Nothing if outcome 2 in right lottery
rdon8	Double or Nothing if outcome 2 in right lottery
rdon9	Double or Nothing if outcome 3 in left lottery
rdon10	Double or Nothing for any outcome in right lottery
rdon11	Double or Nothing if outcome 3 in right lottery
rdon12	Double or Nothing if outcome 2 in right lottery
rdon13	Double or Nothing if outcome 2 in right lottery
rdon14	Double or Nothing if outcome 2 in right lottery
rdon15	Double or Nothing if outcome 2 in right lottery

Appendix B. Additional Results (ONLINE)

Table A5. Summary of demographics.

Variable	Obs	Mean	Std. Dev.	Min	Max
Young	73	0.9726027	0.1643677	0	1
Female	73	0.5479452	0.5011403	0	1
Black	73	0.6712329	0.4730162	0	1
Asian	73	0.1506849	0.3602173	0	1
Business	73	0.3835616	0.4896182	0	1
Freshman	73	0.1780822	0.3852296	0	1
Senior	73	0.3013699	0.4620285	0	1
High GPA	73	0.5616438	0.4996193	0	1
Christian	73	0.7123288	0.4558098	0	1
Insured	73	0.4109589	0.4954127	0	1

Table A6. Expected utility models using simulated choices under correct specifications, individual CRRA coefficients.

Subject	Specification II					Specification IV					Specification V					Specification VI				
	True	Mean	Std. Dev.	89% Cred. Int.		*True*	Mean	Std. Dev.	89% Cred. Int.		*True*	Mean	Std. Dev.	89% Cred. Int.		*True*	Mean	Std. Dev.	89% Cred. Int.	
1	*0.50*	0.444	0.047	0.377	0.520	*0.50*	0.465	0.038	0.401	0.520	*0.65*	0.513	0.096	0.356	0.660	*0.37*	0.382	0.088	0.242	0.524
2	*−0.20*	−0.219	0.054	−0.304	−0.134	*−0.20*	−0.220	0.050	−0.297	−0.137	*0.48*	0.528	0.126	0.341	0.736	*−0.24*	−0.283	0.181	−0.589	−0.014
3	*0.30*	0.322	0.050	0.243	0.405	*0.30*	0.296	0.046	0.226	0.372	*0.67*	0.699	0.083	0.568	0.835	*0.27*	0.280	0.113	0.116	0.467
4	*0.20*	0.201	0.052	0.118	0.284	*0.20*	0.177	0.049	0.103	0.259	*0.18*	0.367 *	0.105	0.194	0.527	*0.10*	0.167	0.135	−0.041	0.382
5	*0.50*	0.452	0.032	0.403	0.507	*0.50*	0.453 *	0.029	0.407	0.499	*0.56*	0.405 *	0.093	0.261	0.555	*0.37*	0.312	0.101	0.153	0.474
6	*0.80*	0.770	0.036	0.711	0.827	*0.80*	0.789	0.032	0.737	0.839	*0.50*	0.628	0.087	0.485	0.757	*0.88*	0.743 *	0.078	0.618	0.866
7	*0.20*	0.145	0.044	0.077	0.217	*0.20*	0.172	0.038	0.108	0.230	*0.41*	0.370	0.117	0.190	0.562	*0.34*	0.178	0.126	−0.014	0.381
8	*0.70*	0.658	0.033	0.609	0.712	*0.70*	0.658	0.028	0.616	0.704	*0.43*	0.526	0.108	0.362	0.707	*0.74*	0.686	0.085	0.546	0.813
9	*0.80*	0.779	0.031	0.728	0.828	*0.80*	0.777	0.027	0.734	0.821	*0.41*	0.358	0.116	0.166	0.533	*0.52*	0.519	0.109	0.352	0.697
10	*0.70*	0.686	0.036	0.627	0.742	*0.70*	0.689	0.033	0.639	0.744	*0.48*	0.403	0.109	0.222	0.566	*0.35*	0.323	0.114	0.144	0.503
11	*0.20*	0.145	0.044	0.077	0.217	*0.20*	0.172	0.038	0.108	0.230	*0.11*	0.331 *	0.124	0.149	0.535	*0.08*	0.165	0.124	−0.028	0.363
12	*0.60*	0.565	0.042	0.498	0.632	*0.60*	0.584	0.038	0.525	0.646	*0.35*	0.390	0.102	0.240	0.553	*0.62*	0.724	0.080	0.602	0.854
13	*0.20*	0.229	0.049	0.149	0.306	*0.20*	0.208	0.046	0.133	0.280	*0.42*	0.457	0.090	0.316	0.603	*0.45*	0.443	0.100	0.289	0.605
14	*−0.30*	−0.312	0.050	−0.391	−0.234	*−0.30*	−0.308	0.047	−0.382	−0.234	*0.56*	0.622	0.078	0.486	0.733	*−0.56*	−0.401	0.157	−0.645	−0.153
15	*0.00*	0.015	0.049	−0.066	0.091	*0.00*	0.016	0.046	−0.058	0.089	*0.50*	0.441	0.107	0.268	0.605	*−0.20*	−0.083	0.163	−0.346	0.168
16	*0.40*	0.351	0.055	0.258	0.432	*0.40*	0.377	0.043	0.308	0.446	*0.42*	0.413	0.091	0.270	0.554	*0.28*	0.371	0.080	0.241	0.499
17	*0.30*	0.238	0.043	0.171	0.309	*0.30*	0.260	0.038	0.202	0.323	*0.62*	0.660	0.095	0.515	0.814	*0.40*	0.429	0.129	0.227	0.636
18	*0.50*	0.452	0.032	0.403	0.507	*0.50*	0.453 *	0.029	0.407	0.499	*0.57*	0.517	0.090	0.367	0.655	*0.65*	0.578	0.089	0.434	0.713
19	*0.00*	−0.004	0.051	−0.085	0.076	*0.00*	−0.027	0.047	−0.105	0.044	*−0.01*	0.437 *	0.100	0.268	0.585	*0.01*	0.260 *	0.147	0.029	0.481
20	*0.40*	0.351	0.055	0.258	0.432	*0.40*	0.377	0.043	0.308	0.446	*0.50*	0.331	0.114	0.156	0.514	*0.48*	0.313 *	0.096	0.155	0.463
21	*0.70*	0.658	0.033	0.609	0.712	*0.70*	0.658	0.028	0.616	0.704	*0.68*	0.658	0.101	0.505	0.819	*0.97*	0.925	0.053	0.855	1.000
22	*−0.10*	−0.097	0.050	−0.182	−0.023	*−0.10*	−0.115	0.047	−0.190	−0.040	*0.36*	0.378	0.107	0.229	0.566	*0.00*	−0.191	0.149	−0.428	0.042
23	*0.00*	0.015	0.049	−0.066	0.091	*0.00*	0.016	0.046	−0.058	0.089	*0.45*	0.339	0.122	0.157	0.544	*−0.02*	−0.072	0.146	−0.304	0.154
24	*0.80*	0.779	0.031	0.728	0.828	*0.80*	0.777	0.027	0.734	0.821	*0.58*	0.464	0.100	0.312	0.626	*1.07*	0.939 *	0.047	0.875	1.000
25	*0.10*	0.108	0.054	0.029	0.200	*0.10*	0.104	0.050	0.025	0.183	*0.52*	0.419	0.103	0.263	0.586	*−0.12*	−0.011	0.159	−0.245	0.247
26	*0.00*	−0.004	0.051	−0.085	0.076	*0.00*	−0.027	0.047	−0.105	0.044	*0.51*	0.345	0.106	0.191	0.520	*−0.37*	−0.112 *	0.154	−0.353	0.138
27	*−0.20*	−0.219	0.054	−0.304	−0.134	*−0.20*	−0.220	0.050	−0.297	−0.137	*0.95*	0.853	0.073	0.751	0.981	*−0.31*	−0.260	0.146	−0.488	−0.028
28	*0.50*	0.528	0.043	0.459	0.595	*0.50*	0.501	0.043	0.434	0.571	*0.21*	0.386	0.121	0.195	0.583	*0.46*	0.469	0.130	0.264	0.675
29	*−0.10*	−0.097	0.050	−0.182	−0.023	*−0.10*	−0.115	0.047	−0.190	−0.040	*0.65*	0.734	0.095	0.590	0.893	*−0.02*	−0.234	0.157	−0.476	0.019
30	*−0.10*	−0.097	0.050	−0.182	−0.023	*−0.10*	−0.115	0.047	−0.190	−0.040	*0.69*	0.513	0.122	0.327	0.710	*0.01*	−0.145	0.170	−0.413	0.129
31	*−0.30*	−0.312	0.050	−0.391	−0.234	*−0.30*	−0.308	0.047	−0.382	−0.234	*0.50*	0.409	0.094	0.257	0.555	*−0.20*	−0.168	0.174	−0.445	0.111
32	*0.50*	0.452	0.032	0.403	0.507	*0.50*	0.453 *	0.029	0.407	0.499	*0.38*	0.302	0.111	0.125	0.475	*0.79*	0.537 *	0.105	0.375	0.704
33	*0.40*	0.435	0.050	0.356	0.515	*0.40*	0.413	0.048	0.338	0.493	*0.49*	0.447	0.112	0.261	0.618	*0.48*	0.393	0.130	0.184	0.596
34	*0.20*	0.145	0.044	0.077	0.217	*0.20*	0.172	0.038	0.108	0.230	*0.59*	0.492	0.107	0.327	0.667	*0.23*	0.269	0.116	0.084	0.445
35	*0.70*	0.686	0.036	0.627	0.742	*0.70*	0.689	0.033	0.639	0.744	*0.44*	0.461	0.091	0.321	0.610	*0.55*	0.495	0.088	0.360	0.636
36	*0.70*	0.658	0.033	0.609	0.712	*0.70*	0.658	0.028	0.616	0.704	*0.51*	0.493	0.092	0.347	0.636	*0.48*	0.534	0.083	0.405	0.664
37	*0.30*	0.314	0.050	0.240	0.397	*0.30*	0.309	0.049	0.231	0.386	*0.51*	0.361	0.131	0.150	0.563	*0.20*	0.071	0.160	−0.175	0.331
38	*0.70*	0.677	0.045	0.607	0.752	*0.70*	0.701	0.038	0.639	0.758	*0.38*	0.417	0.115	0.235	0.600	*0.61*	0.625	0.083	0.490	0.753
39	*0.40*	0.359	0.038	0.299	0.418	*0.40*	0.365	0.035	0.310	0.422	*0.53*	0.490	0.118	0.305	0.673	*0.34*	0.179	0.123	−0.017	0.373
40	*0.20*	0.145	0.044	0.077	0.217	*0.20*	0.172	0.038	0.108	0.230	*0.70*	0.631	0.094	0.486	0.780	*−0.04*	0.184 *	0.107	0.012	0.347

Table A6. *Cont.*

Subject	Specification II					Specification IV					Specification V					Specification VI				
	True	Mean		Std. Dev.	89% Cred. Int.	*True*	Mean		Std. Dev.	89% Cred. Int.	*True*	Mean		Std. Dev.	89% Cred. Int.	*True*	Mean		Std. Dev.	89% Cred. Int.
41	*0.20*	0.201		0.052	0.118 0.284	*0.20*	0.177		0.049	0.103 0.259	*0.77*	0.529	*	0.109	0.358 0.696	*0.24*	0.204		0.160	−0.056 0.448
42	*0.00*	0.015		0.049	−0.066 0.091	*0.00*	0.016		0.046	−0.058 0.089	*0.34*	0.374		0.107	0.207 0.543	*0.05*	0.124		0.160	−0.134 0.374
43	*0.80*	0.779		0.031	0.728 0.828	*0.80*	0.777		0.027	0.734 0.821	*0.75*	0.811		0.083	0.684 0.952	*0.76*	0.916	*	0.062	0.833 1.000
44	*0.70*	0.686		0.036	0.627 0.742	*0.70*	0.689		0.033	0.639 0.744	*0.56*	0.609		0.086	0.487 0.756	*0.83*	0.668	*	0.082	0.541 0.800
45	*0.50*	0.452		0.032	0.403 0.507	*0.50*	0.453	*	0.029	0.407 0.499	*0.56*	0.562		0.079	0.442 0.695	*0.55*	0.565		0.075	0.447 0.684
46	*0.70*	0.677		0.045	0.607 0.752	*0.70*	0.701		0.038	0.639 0.758	*0.52*	0.584		0.107	0.419 0.757	*0.91*	0.798		0.084	0.662 0.931
47	*0.60*	0.565		0.042	0.498 0.632	*0.60*	0.584		0.038	0.525 0.646	*0.74*	0.654		0.079	0.531 0.781	*0.62*	0.565		0.087	0.423 0.700
48	*0.30*	0.238		0.043	0.171 0.309	*0.30*	0.260		0.038	0.202 0.323	*0.53*	0.494		0.100	0.339 0.659	*0.51*	0.362		0.112	0.180 0.536
49	*0.80*	0.770		0.036	0.711 0.827	*0.80*	0.789		0.032	0.737 0.839	*0.86*	0.749		0.089	0.624 0.909	*1.09*	0.935	*	0.049	0.869 1.000
50	*0.20*	0.221		0.053	0.139 0.305	*0.20*	0.221		0.049	0.138 0.295	*0.23*	0.294		0.125	0.087 0.481	*0.27*	0.133		0.141	−0.089 0.360
51	*0.50*	0.452		0.032	0.403 0.507	*0.50*	0.453	*	0.029	0.407 0.499	*0.40*	0.483		0.102	0.321 0.646	*0.57*	0.477		0.105	0.310 0.640
52	*0.40*	0.359		0.038	0.299 0.418	*0.40*	0.365		0.035	0.310 0.422	*0.05*	0.211		0.109	0.036 0.381	*0.29*	0.293		0.121	0.112 0.490
53	*0.70*	0.658		0.033	0.609 0.712	*0.70*	0.658		0.028	0.616 0.704	*0.44*	0.380		0.098	0.225 0.535	*0.64*	0.541		0.088	0.400 0.677
54	*0.00*	−0.004		0.051	−0.085 0.076	*0.00*	−0.027		0.047	−0.105 0.044	*0.75*	0.699		0.078	0.574 0.823	*0.33*	0.280		0.149	0.053 0.502
55	*0.50*	0.452		0.032	0.403 0.507	*0.50*	0.453	*	0.029	0.407 0.499	*0.61*	0.459		0.108	0.298 0.642	*0.67*	0.452	*	0.115	0.274 0.637
56	*0.70*	0.686		0.036	0.627 0.742	*0.70*	0.689		0.033	0.639 0.744	*0.47*	0.390		0.103	0.231 0.559	*0.89*	0.757		0.084	0.625 0.893
57	*−0.30*	−0.312		0.050	−0.391 −0.234	*−0.30*	−0.308		0.047	−0.382 −0.234	*0.48*	0.489		0.099	0.345 0.655	*0.19*	−0.187	*	0.182	−0.470 0.106
58	*−0.20*	−0.219		0.054	−0.304 −0.134	*−0.20*	−0.220		0.050	−0.297 −0.137	*0.68*	0.693		0.108	0.522 0.866	*−0.22*	−0.369		0.166	−0.612 −0.093
59	*0.30*	0.314		0.050	0.240 0.397	*0.30*	0.309		0.049	0.231 0.386	*0.25*	0.508	*	0.091	0.364 0.648	*0.16*	0.419	*	0.110	0.242 0.587
60	*0.20*	0.221		0.053	0.139 0.305	*0.20*	0.221		0.049	0.138 0.295	*0.26*	0.396		0.094	0.236 0.531	*0.11*	0.258		0.115	0.073 0.435
61	*0.40*	0.351		0.055	0.258 0.432	*0.40*	0.377		0.043	0.308 0.446	*0.63*	0.493		0.104	0.337 0.662	*0.19*	0.256		0.099	0.100 0.419
62	*1.00*	0.985	*	0.014	0.967 1.000	*1.00*	0.982	*	0.016	0.961 1.000	*0.47*	0.501		0.095	0.353 0.655	*1.02*	0.944	*	0.045	0.883 1.000
63	*0.70*	0.658		0.033	0.609 0.712	*0.70*	0.658		0.028	0.616 0.704	*0.58*	0.517		0.102	0.368 0.686	*1.03*	0.916	*	0.057	0.839 1.000
64	*0.50*	0.452		0.032	0.403 0.507	*0.50*	0.453	*	0.029	0.407 0.499	*0.44*	0.493		0.126	0.304 0.702	*0.68*	0.560		0.135	0.349 0.775
65	*0.30*	0.238		0.043	0.171 0.309	*0.30*	0.260		0.038	0.202 0.323	*0.39*	0.325		0.122	0.130 0.520	*0.34*	0.221		0.129	0.020 0.426
66	*0.50*	0.528		0.043	0.459 0.595	*0.50*	0.501		0.043	0.434 0.571	*0.57*	0.538		0.120	0.360 0.738	*0.46*	0.450		0.149	0.218 0.687
67	*0.50*	0.452		0.032	0.403 0.507	*0.50*	0.453	*	0.029	0.407 0.499	*0.44*	0.385		0.110	0.220 0.570	*0.25*	0.315		0.121	0.124 0.504
68	*0.90*	0.892		0.036	0.832 0.945	*0.90*	0.894		0.032	0.843 0.943	*0.47*	0.379		0.103	0.224 0.548	*0.86*	0.924		0.055	0.851 1.000
69	*0.00*	−0.004		0.051	−0.085 0.076	*0.00*	−0.027		0.047	−0.105 0.044	*0.50*	0.399		0.104	0.229 0.556	*−0.08*	−0.166		0.141	−0.381 0.061
70	*0.60*	0.565		0.042	0.498 0.632	*0.60*	0.584		0.038	0.525 0.646	*0.79*	0.748		0.083	0.626 0.885	*0.40*	0.528		0.097	0.370 0.678
71	*0.70*	0.686		0.036	0.627 0.742	*0.70*	0.689		0.033	0.639 0.744	*0.71*	0.585	*	0.081	0.448 0.703	*0.46*	0.560		0.090	0.422 0.702
72	*0.40*	0.359		0.038	0.299 0.418	*0.40*	0.365		0.035	0.310 0.422	*0.67*	0.716		0.078	0.594 0.842	*0.74*	0.720		0.078	0.605 0.849
73	*0.20*	0.145		0.044	0.077 0.217	*0.20*	0.172		0.038	0.108 0.230	*0.52*	0.526		0.119	0.351 0.723	*0.46*	0.240		0.141	0.023 0.474
Average	*0.36*	0.341				*0.37*	0.344				*0.51*	0.493				*0.37*	0.345			
Std. Dev.	*0.32*	0.318				*0.37*	0.321				*0.18*	0.134				*0.37*	0.346			

* Indicates that the true parameter is outside of the 89% credible interval of the posterior sample.

Table A7. Expected utility models using simulated dataset VI under mis-specified specifications, individual CRRA coefficients.

Subject	True Value	Specification II					Specification IV					Specification V					Specification VI				
		Mean		Std. Dev.	89% Cred. Int.		Mean		Std. Dev.	89% Cred. Int.		Mean		Std. Dev.	89% Cred. Int.		Mean		Std. Dev.	89% Cred. Int.	
1	0.37	0.546	*	0.037	0.488	0.602	0.539	*	0.038	0.481	0.598	0.313		0.118	0.134	0.504	0.382		0.088	0.242	0.524
2	−0.24	−0.170		0.047	−0.248	−0.097	−0.160	*	0.047	−0.238	−0.087	−0.153		0.229	−0.491	0.207	−0.283		0.181	−0.589	−0.014
3	0.27	0.268		0.044	0.196	0.337	0.290		0.047	0.217	0.367	0.281		0.115	0.104	0.464	0.280		0.113	0.116	0.467
4	0.10	0.207	*	0.047	0.131	0.280	0.215	*	0.049	0.134	0.289	0.182		0.153	−0.065	0.411	0.167		0.135	−0.041	0.382
5	0.37	0.509	*	0.029	0.464	0.555	0.522	*	0.030	0.477	0.573	0.251		0.127	0.043	0.445	0.312		0.101	0.153	0.474
6	0.88	0.815	*	0.031	0.768	0.866	0.809	*	0.031	0.761	0.859	0.712	*	0.081	0.587	0.843	0.743	*	0.078	0.618	0.866
7	0.34	0.212	*	0.035	0.156	0.266	0.199	*	0.038	0.138	0.259	0.211		0.172	−0.055	0.477	0.178		0.126	−0.014	0.381
8	0.74	0.716		0.028	0.673	0.763	0.718		0.028	0.670	0.761	0.664		0.093	0.522	0.814	0.686		0.085	0.546	0.813
9	0.52	0.778	*	0.027	0.733	0.820	0.793	*	0.027	0.753	0.839	0.405		0.137	0.186	0.609	0.519		0.109	0.352	0.697
10	0.35	0.650	*	0.032	0.599	0.702	0.649	*	0.035	0.591	0.702	0.165		0.145	−0.058	0.390	0.323		0.114	0.144	0.503
11	0.08	0.212	*	0.035	0.156	0.266	0.199	*	0.038	0.138	0.259	0.181		0.187	−0.106	0.475	0.165		0.124	−0.028	0.363
12	0.62	0.608		0.039	0.548	0.670	0.614		0.036	0.555	0.672	0.726		0.084	0.590	0.857	0.724		0.080	0.602	0.854
13	0.45	0.141	*	0.045	0.071	0.215	0.146	*	0.050	0.068	0.229	0.487		0.093	0.340	0.635	0.443		0.100	0.289	0.605
14	−0.56	−0.298	*	0.044	−0.368	−0.228	−0.304	*	0.047	−0.377	−0.228	−0.297		0.198	−0.613	0.005	−0.401		0.157	−0.645	−0.153
15	−0.20	−0.029	*	0.045	−0.099	0.042	−0.034	*	0.051	−0.113	0.050	−0.040		0.201	−0.349	0.286	−0.083		0.163	−0.346	0.168
16	0.28	0.419	*	0.042	0.350	0.485	0.395	*	0.043	0.327	0.460	0.372		0.107	0.222	0.552	0.371		0.080	0.241	0.499
17	0.40	0.339	*	0.037	0.278	0.395	0.343		0.036	0.288	0.400	0.510		0.144	0.294	0.734	0.429		0.129	0.227	0.636
18	0.65	0.509	*	0.029	0.464	0.555	0.522	*	0.030	0.477	0.573	0.578		0.098	0.422	0.728	0.578		0.089	0.434	0.713
19	0.01	−0.001		0.045	−0.074	0.069	0.019		0.046	−0.055	0.093	0.389	*	0.125	0.210	0.597	0.260	*	0.147	0.029	0.481
20	0.48	0.419		0.042	0.350	0.485	0.395	*	0.043	0.327	0.460	0.185	*	0.176	−0.075	0.464	0.313	*	0.096	0.155	0.463
21	0.97	0.716	*	0.028	0.673	0.763	0.718	*	0.028	0.670	0.761	0.937		0.050	0.871	1.000	0.925		0.053	0.855	1.000
22	0.00	−0.128	*	0.047	−0.205	−0.055	−0.125	*	0.048	−0.202	−0.047	−0.118		0.169	−0.391	0.130	−0.191		0.149	−0.428	0.042
23	−0.02	−0.029		0.045	−0.099	0.042	−0.034		0.051	−0.113	0.050	−0.033		0.175	−0.290	0.253	−0.072		0.146	−0.304	0.154
24	1.07	0.778	*	0.027	0.733	0.820	0.793	*	0.027	0.753	0.839	0.933	*	0.052	0.863	1.000	0.939	*	0.047	0.875	1.000
25	−0.12	0.099	*	0.049	0.020	0.174	0.110	*	0.050	0.036	0.193	−0.002		0.179	−0.269	0.282	−0.011		0.159	−0.245	0.247
26	−0.37	−0.001	*	0.045	−0.074	0.069	0.019	*	0.046	−0.055	0.093	−0.081	*	0.184	−0.346	0.224	−0.112	*	0.154	−0.353	0.138
27	−0.31	−0.170	*	0.047	−0.248	−0.097	−0.160	*	0.047	−0.238	−0.087	−0.184		0.163	−0.424	0.082	−0.260		0.146	−0.488	−0.028
28	0.46	0.476		0.039	0.414	0.537	0.486		0.044	0.416	0.557	0.429		0.161	0.196	0.690	0.469		0.130	0.264	0.675
29	−0.02	−0.128	*	0.047	−0.205	−0.055	−0.125	*	0.048	−0.202	−0.047	−0.195		0.192	−0.497	0.100	−0.234		0.157	−0.476	0.019
30	0.01	−0.128	*	0.047	−0.205	−0.055	−0.125	*	0.048	−0.202	−0.047	−0.017		0.204	−0.306	0.325	−0.145		0.170	−0.413	0.129
31	−0.20	−0.298	*	0.044	−0.368	−0.228	−0.304	*	0.047	−0.377	−0.228	0.150	*	0.147	−0.087	0.368	−0.168		0.174	−0.445	0.111
32	0.79	0.509	*	0.029	0.464	0.555	0.522	*	0.030	0.477	0.573	0.539	*	0.118	0.362	0.734	0.537	*	0.105	0.375	0.704
33	0.48	0.348	*	0.047	0.274	0.423	0.342	*	0.050	0.261	0.421	0.389		0.163	0.136	0.639	0.393		0.130	0.184	0.596
34	0.23	0.212		0.035	0.156	0.266	0.199		0.038	0.138	0.259	0.355		0.129	0.152	0.547	0.269		0.116	0.084	0.445
35	0.55	0.650	*	0.032	0.599	0.702	0.649	*	0.035	0.591	0.702	0.440		0.102	0.288	0.602	0.495		0.088	0.360	0.636
36	0.48	0.716	*	0.028	0.673	0.763	0.718	*	0.028	0.670	0.761	0.478		0.101	0.339	0.650	0.534		0.083	0.405	0.664
37	0.20	0.306	*	0.045	0.238	0.381	0.306	*	0.050	0.229	0.389	−0.070		0.212	−0.410	0.255	0.071		0.160	−0.175	0.331
38	0.61	0.688	*	0.036	0.631	0.745	0.665		0.039	0.608	0.729	0.577		0.102	0.432	0.754	0.625		0.083	0.490	0.753
39	0.34	0.381		0.035	0.325	0.435	0.378		0.035	0.320	0.434	0.032	*	0.181	−0.267	0.292	0.179		0.123	−0.017	0.373
40	−0.04	0.212	*	0.035	0.156	0.266	0.199	*	0.038	0.138	0.259	0.210	*	0.128	0.023	0.419	0.184	*	0.107	0.012	0.347

Table A7. *Cont.*

Subject	True Value	Specification II					Specification IV					Specification V					Specification VI				
		Mean		Std. Dev.	89% Cred. Int.		Mean		Std. Dev.	89% Cred. Int.		Mean		Std. Dev.	89% Cred. Int.		Mean		Std. Dev.	89% Cred. Int.	
41	0.24	0.207		0.047	0.131	0.280	0.215		0.049	0.134	0.289	0.243		0.196	−0.042	0.558	0.204		0.160	−0.056	0.448
42	0.05	−0.029	*	0.045	−0.099	0.042	−0.034	*	0.051	−0.113	0.050	0.276		0.157	0.039	0.517	0.124		0.160	−0.134	0.374
43	0.76	0.778		0.027	0.733	0.820	0.793		0.027	0.753	0.839	0.908	*	0.064	0.824	1.000	0.916	*	0.062	0.833	1.000
44	0.83	0.650	*	0.032	0.599	0.702	0.649	*	0.035	0.591	0.702	0.648	*	0.084	0.519	0.783	0.668	*	0.082	0.541	0.800
45	0.55	0.509		0.029	0.464	0.555	0.522		0.030	0.477	0.573	0.562		0.079	0.446	0.693	0.565		0.075	0.447	0.684
46	0.91	0.688	*	0.036	0.631	0.745	0.665	*	0.039	0.608	0.729	0.825		0.091	0.689	0.978	0.798		0.084	0.662	0.931
47	0.62	0.608		0.039	0.548	0.670	0.614		0.036	0.555	0.672	0.541		0.098	0.394	0.699	0.565		0.087	0.423	0.700
48	0.51	0.339	*	0.037	0.278	0.395	0.343	*	0.036	0.288	0.400	0.389		0.130	0.183	0.582	0.362		0.112	0.180	0.536
49	1.09	0.815	*	0.031	0.768	0.866	0.809	*	0.031	0.761	0.859	0.933	*	0.052	0.862	1.000	0.935	*	0.049	0.869	1.000
50	0.27	0.178	*	0.047	0.103	0.250	0.162	*	0.053	0.081	0.251	0.126		0.200	−0.169	0.450	0.133		0.141	−0.089	0.360
51	0.57	0.509	*	0.029	0.464	0.555	0.522		0.030	0.477	0.573	0.448		0.125	0.248	0.634	0.477		0.105	0.310	0.640
52	0.29	0.381	*	0.035	0.325	0.435	0.378	*	0.035	0.320	0.434	0.224		0.168	−0.042	0.479	0.293		0.121	0.112	0.490
53	0.64	0.716	*	0.028	0.673	0.763	0.718	*	0.028	0.670	0.761	0.481		0.110	0.316	0.657	0.541		0.088	0.400	0.677
54	0.33	−0.001	*	0.045	−0.074	0.069	0.019	*	0.046	−0.055	0.093	0.407		0.118	0.222	0.580	0.280		0.149	0.053	0.502
55	0.67	0.509	*	0.029	0.464	0.555	0.522	*	0.030	0.477	0.573	0.413	*	0.145	0.197	0.635	0.452	*	0.115	0.274	0.637
56	0.89	0.650	*	0.032	0.599	0.702	0.649	*	0.035	0.591	0.702	0.748		0.089	0.610	0.894	0.757		0.084	0.625	0.893
57	0.19	−0.298	*	0.044	−0.368	−0.228	−0.304	*	0.047	−0.377	−0.228	0.188		0.166	−0.048	0.469	−0.187	*	0.182	−0.470	0.106
58	−0.22	−0.170		0.047	−0.248	−0.097	−0.160		0.047	−0.238	−0.087	−0.323		0.201	−0.668	−0.033	−0.369		0.166	−0.612	−0.093
59	0.16	0.306	*	0.045	0.238	0.381	0.306	*	0.050	0.229	0.389	0.452	*	0.110	0.280	0.614	0.419	*	0.110	0.242	0.587
60	0.11	0.178		0.047	0.103	0.250	0.162		0.053	0.081	0.251	0.300		0.127	0.105	0.500	0.258		0.115	0.073	0.435
61	0.19	0.419	*	0.042	0.350	0.485	0.395	*	0.043	0.327	0.460	0.012		0.181	−0.273	0.285	0.256		0.099	0.100	0.419
62	1.02	0.985	*	0.014	0.968	1.000	0.989	*	0.010	0.975	1.000	0.927	*	0.053	0.856	1.000	0.944	*	0.045	0.883	1.000
63	1.03	0.716	*	0.028	0.673	0.763	0.718	*	0.028	0.670	0.761	0.927	*	0.054	0.853	1.000	0.916	*	0.057	0.839	1.000
64	0.68	0.509	*	0.029	0.464	0.555	0.522	*	0.030	0.477	0.573	0.561		0.174	0.282	0.825	0.560		0.135	0.349	0.775
65	0.34	0.339		0.037	0.278	0.395	0.343		0.036	0.288	0.400	0.162		0.186	−0.133	0.439	0.221		0.129	0.020	0.426
66	0.46	0.476		0.039	0.414	0.537	0.486		0.044	0.416	0.557	0.396		0.178	0.122	0.682	0.450		0.149	0.218	0.687
67	0.25	0.509	*	0.029	0.464	0.555	0.522	*	0.030	0.477	0.573	0.219		0.153	−0.006	0.463	0.315		0.121	0.124	0.504
68	0.86	0.857		0.032	0.808	0.910	0.845		0.029	0.797	0.892	0.914		0.061	0.832	1.000	0.924		0.055	0.851	1.000
69	−0.08	−0.001	*	0.045	−0.074	0.069	0.019	*	0.046	−0.055	0.093	−0.155		0.157	−0.400	0.080	−0.166		0.141	−0.381	0.061
70	0.40	0.608	*	0.039	0.548	0.670	0.614	*	0.036	0.555	0.672	0.486		0.116	0.302	0.660	0.528		0.097	0.370	0.678
71	0.46	0.650	*	0.032	0.599	0.702	0.649	*	0.035	0.591	0.702	0.510		0.102	0.350	0.665	0.560		0.090	0.422	0.702
72	0.74	0.381	*	0.035	0.325	0.435	0.378	*	0.035	0.320	0.434	0.781		0.082	0.657	0.920	0.720		0.078	0.605	0.849
73	0.46	0.212	*	0.035	0.156	0.266	0.199	*	0.038	0.138	0.259	0.387		0.201	0.079	0.696	0.240		0.141	0.023	0.474
Average	0.37			0.359					0.361					0.353					0.345		
Std. Dev.	0.37			0.322					0.322					0.322					0.346		

* Indicates that the true parameter is outside of the 89% credible interval of the posterior sample.

Appendix C. MCMC Convergence Diagnostics (ONLINE)

This appendix documents the convergence behavior of the Bayesian hierarchical models reported in Tables 2–5. All models were estimated in Stata 16.1 using *bayesmh*, combining an adaptive random-walk Metropolis–Hastings sampler for the risk-preference parameters with Gibbs updates for the conjugate hierarchical variance components. Each model was run for 10,000 posterior draws following a 2500-iteration burn-in (12,500 iterations in total), using a fixed random seed for reproducibility. Observed acceptance rates ranged from approximately 0.43 to 0.52.

Appendix C.1. Diagnostics

We assess convergence parameter by parameter using three complementary diagnostics. First, for each parameter we compute the Monte Carlo standard error (MCSE). Because *bayesmh* computes the MCSE as the posterior standard deviation divided by the square root of the effective sample size, the effective sample size (ESS) for each reported parameter is recovered as ESS = $(\mathrm{SD}/\mathrm{MCSE})^2$. Second, we compute the rank-normalized split-$\hat{R}$ statistic (Vehtari et al. 2021), a modern refinement of the Gelman–Rubin potential scale reduction factor (Gelman and Rubin 1992; Brooks and Gelman 1998). Because a single chain was run per model, split-$\hat{R}$ is obtained by dividing each chain into two halves and comparing within- and between-half variances, following the single-chain extension of Vats and Knudson (2021). Third, we examine trace, autocorrelation, and split-half kernel-density plots for each parameter.

As reference thresholds we treat an effective sample size of 100 as the minimum for a stable posterior mean and 400 as the standard for reliable bulk and tail summaries (Vehtari et al. 2021). For the potential scale reduction factor we adopt the conventional convergence criterion of 1.2 (Brooks and Gelman 1998, p. 444), noting 1.1 (Gelman et al. 2013) and 1.01 (Vehtari et al. 2021; Vats and Knudson 2021) as progressively stricter benchmarks; credible-interval endpoints are the most demanding quantities and benefit from substantially larger effective samples (Kruschke 2015).

Appendix C.2. Convergence: Split-$\hat{R}$

Table A8 reports, for each model, the maximum split-$\hat{R}$ across its reported parameters and the number exceeding the conventional criterion of 1.2 (Brooks and Gelman 1998, p. 444) and the stricter 1.1 benchmark.

Table A8. Split-$\hat{R}$ summary by model.

Model	n	Max Split-$\hat{R}$	# $\hat{R}$ > 1.2	# $\hat{R}$ > 1.1
Expected Utility—simulated, correct (Table 2)	33	1.05	0	0
Expected Utility—simulated, mis-specified (Table 3)	33	1.05	0	0
Expected Utility—observed (Table 4)	33	1.02	0	0
Rank-Dependent Utility—observed (Table 5)	79	1.29	3	8

All three EUT models are converged by any standard: the maximum split-$\hat{R}$ is 1.05 or below (1.017 for the observed-data model), so no single parameter approaches even the stricter 1.1 benchmark, let alone the 1.2 criterion. This holds for the observed-data model (Table 4) whose specification VI contains the lowest-ESS EUT parameters. Under EUT the low effective sample sizes therefore reflect slow mixing (high autocorrelation) rather than any failure to reach the stationary distribution.

The RDU model (Table 5) is an exception. Three parameters exceed the 1.2 criterion and eight exceed the stricter 1.1 benchmark; all are r and η demographic coefficients in specifications II, IV, and VI, together with the φ and μ coefficients in specification VI. The

three that exceed 1.2 are the η constant in specification VI (split-$\hat{R}$= 1.29), r × Black in specification IV (1.23), and η × Black in specification IV (1.22).

Appendix C.3. Effective Sample Sizes by Specification

Tables A9–A12 summarize the effective sample sizes of the reported (population-level) parameters within each specification of each model. Here, *n* is the number of reported parameters in the specification, and the final column counts those with an effective sample size below 100. Individual-level parameters (subject-specific r and μ), which mix substantially faster, are not included in these counts.

Table A9. Effective sample size: EUT, simulated data, correct specifications (Table 2).

Spec	N	Min	Max	Avg	Params ESS < 100
I	2	1895	2443	2169	0
II	7	47	1267	331	2
III	3	85	1918	854	1
IV	8	81	780	278	1
V	4	200	2074	913	0
VI	9	19	306	106	5

Table A10. Effective sample size: EUT, simulated dataset VI under mis-specification (Table 3).

Spec	n	Min	Max	Avg	Params ESS < 100
I	2	2001	2305	2153	0
II	7	48	1515	505	1
III	3	410	648	541	0
IV	8	29	942	326	4
V	4	453	3469	1654	0
VI	9	19	306	106	5

Table A11. Effective sample size: EUT, observed choices (Table 4).

Spec	n	Min	Max	Avg	Params ESS < 100
I	2	1640	2090	1865	0
II	7	187	1639	713	0
III	3	439	668	580	0
IV	8	122	968	467	0
V	4	278	5624	2111	0
VI	9	17	254	83	6

Table A12. Effective sample size: RDU, observed choices (Table 5).

Spec	n	Min	Max	Avg	Params ESS < 100
I	4	153	1125	558	0
II	19	20	546	151	11
III	5	174	644	452	0
IV	20	15	260	80	14
V	8	52	521	232	1
VI	23	18	202	47	22

The pooled specification (I) and the hierarchical specifications without demographic structure (III and V) mix well in every model, with effective sample sizes in the hundreds to several thousand, and essentially no parameters below 100. Low effective sample sizes

are confined to the demographic-coefficient blocks of the most heavily parameterized specifications (IV and VI), and are markedly more prevalent under RDU (48 of 79 reported parameters below 100) than under EUT (6 of 33 in the observed-data model).

Appendix C.4. Graphical Diagnostics

The trace, autocorrelation, and split-half density plots are consistent with these statistics. The individual-level parameters, the noise term μ and its variance, and the random-coefficient variances mix well throughout. We observe stationary-noise traces, autocorrelation decaying to zero within roughly 5–15 lags, and coincident half-chain densities. The demographic-mean coefficients of the richest specifications are highly autocorrelated in both models, which accounts for their low effective sample sizes, but the two half-chain densities coincide for the expected-utility models, matching their split-$\hat{R}$ of 1.05 or below, and separate only for the flagged rank-dependent coefficients, matching their elevated split-$\hat{R}$. In every case the traces are stationary, fluctuating around a stable level without drift, and the marginal posteriors are unimodal, so the posterior means and signs are reliable.

A further consideration substantially mitigates the slow mixing of these demographic coefficients. In the hierarchical specifications the objects of practical interest are the individual-level preference parameters: each subject's r, η, and φ, reconstructed from that subject's demographics, the sampled coefficients, and the subject-specific random effect. These subject-level parameters mix well even in specification VI, where the population coefficients mix most slowly: their autocorrelation decays to roughly 0.3 within ten to fifteen lags and their first- and second-half posterior densities essentially coincide. The diagnostic figures at the end of this appendix show this pattern using φ parameters in specification VI as an example. The population demographic coefficients, by contrast, are auxiliary parameters: they parameterize the prior through which the individual estimates are inferred, rather than being quantities of direct interest in their own right. Their slow mixing reflects the weak identification (trade-off) among them, but because the data pin down each subject's preference, the linear combinations that constitute the individual parameters are well determined. Convergence of the quantities actually used in practice is therefore not in question. In this sense, the distinction between the two is one of interest *versus* non-interest for the question being asked, not a feature of the model itself (Gao et al. 2023).

Appendix C.5. Interpretation and Robustness

The convergence concern is thus isolated to a small set of demographic-mean coefficients in the rank-dependent specifications. This localization is consistent with weak identification, and strong posterior correlation, among the curvature parameter (r) and the probability-weighting parameters (η and φ). This is a feature of the RDU model that is absent from the EUT model, rather than with any deficiency of the sampler: estimated with the identical sampler and settings, all EUT models converge.

For the affected RDU coefficients (split-$\hat{R} \leq 1.29$), 10,000 draws have not fully resolved the posterior tails. The posterior means and signs are nonetheless stable across the chain halves, so it is the width of the credible intervals, not the location of these coefficients, that is imprecisely estimated. For the quantities of practical interest, the individual-level parameters and every EUT estimate, the Gelman–Rubin ratios remain below threshold, and the first- and second-half densities overlap, so the modest sampling efficiency there costs precision rather than validity (as in Gao et al. 2023).

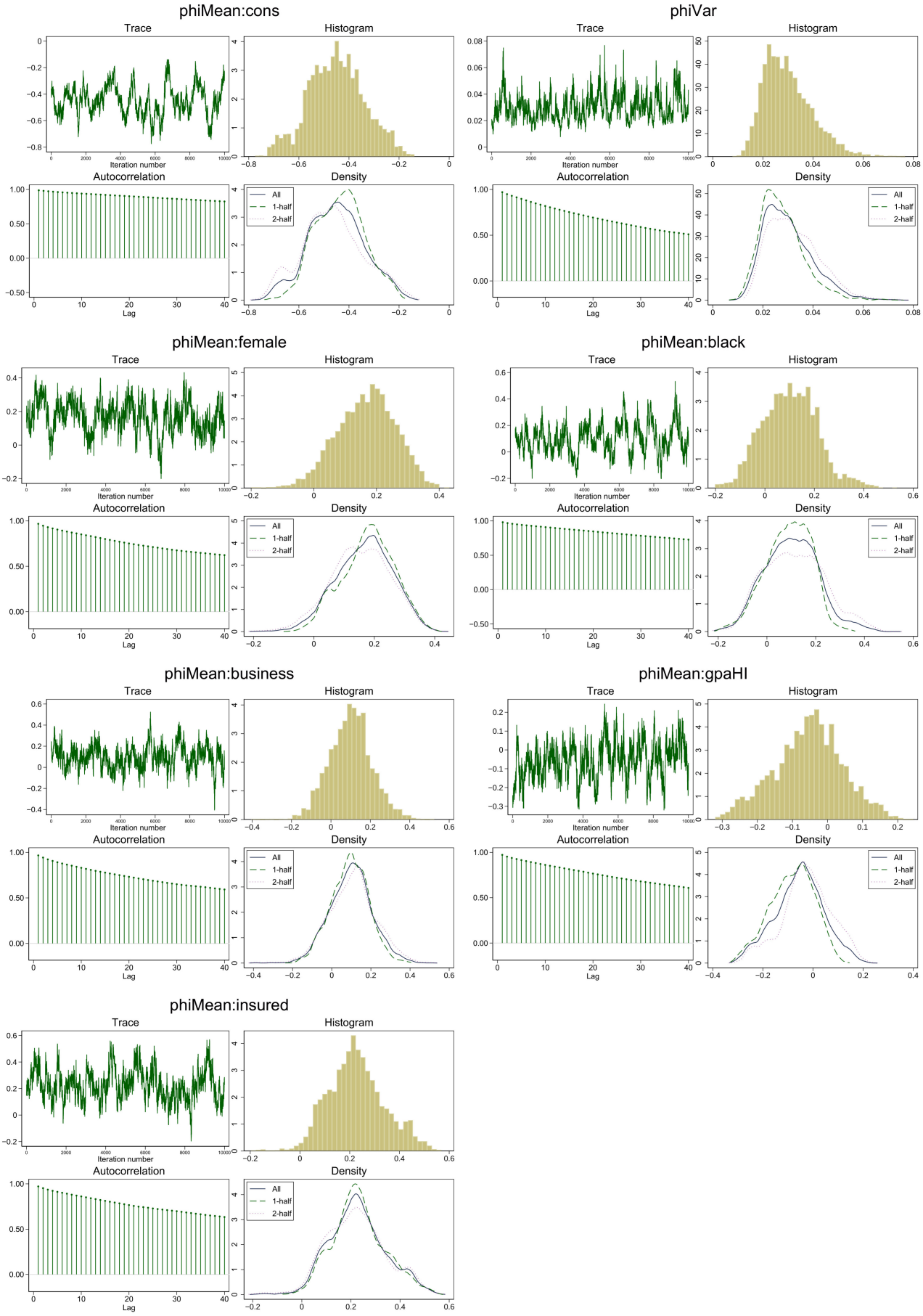


Figure A1. Diagnostic plots of demographic coefficients of φ in RDU specification VI.

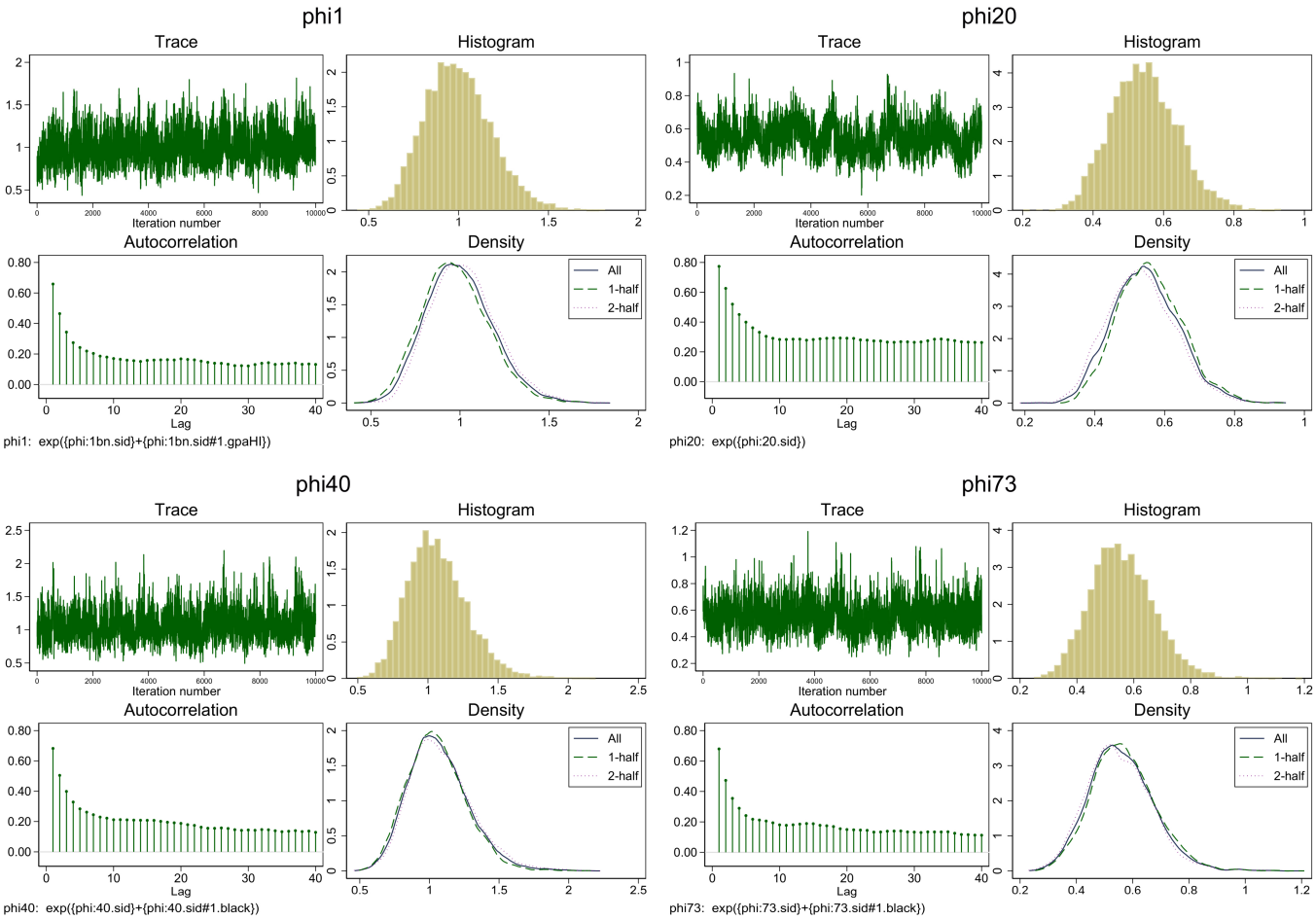


Figure A2. Diagnostic plots of individual φ of subjects 1, 20, 40 and 73 in RDU specification VI.

Notes

1 That is, of course, the standard probabilistic foundation laid out in any modern treatment of Bayesian statistics; see Gelman et al. (2013, Chapter 5) or Kruschke (2015, Chapter 9). The substantive point is that exchangeability is not just a convenience: it is the assumption that justifies the entire pooling apparatus.

2 Hierarchical Bayesian inference of individual preferences has a long tradition in marketing, going back to Rossi and Allenby (1993) and Allenby and Rossi (1999), and has been used to estimate Cumulative Prospect Theory parameters by Nilsson et al. (2011) and to estimate risk preferences for individual behavioral welfare evaluation by Gao et al. (2023).

3 While we choose to model σ_r^2 as homogeneous across different demographic groups here, a potential extension is to also allow this parameter to vary with observables.

4 Several cautions are worth bearing in mind, and we follow Harrison and Rutström (2009) and the wider Bayesian methodological literature in treating Bayes factors with some care. First, the inferential task here is evaluation, not selection in the nihilistic sense of rejecting a model for all purposes. "Consistency with the data" can mean in-sample prediction, out-of-sample prediction, or prediction conditional on theoretical priors, and the various criteria need not agree; until they do, no single metric should be used dogmatically. Second, hypothesis tests that depend on a specific structural form, such as the normative welfare evaluation in Harrison and Ng (2016) and Gao et al. (2023), which requires a utility function to compute certainty equivalents, cannot be allowed to depend on a horse race that might eliminate utility-based models in favor of something atheoretical. This is what Leamer (1978, Chapter 6) called a causally-constrained specification search: the prior that a utility function is in place is part of the inference, not a parameter for the data to overturn. Third, Bayes factors are sensitive to the priors assumed for each model. Following Vanpaemel (2010), we view that sensitivity as appropriate behavior for a Bayesian comparison, not a defect. Fourth, computing marginal likelihoods for large hierarchical models is genuinely difficult, and we follow Gronau et al. (2017) in being cautious about the numerical reliability of the Bayes factor estimates themselves, particularly when bridge sampling or related techniques would be needed for full precision. Throughout, then, Bayes factors are one summary of how well the alternative amounts of conditioning describe the observed choices in our sample under the priors we have assumed. They are useful for that purpose, but not a verdict.

[5] We do not allow demographic differences to enter the prior for μ_i for three reasons. First, μ_i is a nuisance parameter that captures the noisiness of decision making in this specific binary-choice task, not a preference. Second, we have no *a priori* reason to expect this decision noise to differ systematically across demographic groups in our subject pool. Third, allowing demographic structure in μ_i would substantially complicate the parameter space without correspondingly clear identification benefits.

[6] We choose 89% deliberately, following McElreath (2020, p. 56), to avoid confusion with popular levels of statistical significance in non-Bayesian analyses.

[7] A natural question to ask is whether one really needs covariates in the hyper-parameters at all. Post-stratification of the estimated individual-level distribution from the experimental sub-sample, by the demographic margins of the wider population, would deliver a population-level distribution of risk preferences without requiring any conditioning at the prior stage. For some inferential purposes, most obviously, computing a population-weighted mean or quantile of risk aversion for a cost–benefit exercise, that may be all one needs, and the comparison with the unconditional hierarchical estimates we reported suggests that it would often be close to adequate in practice. For other purposes it is not enough. If the analyst wants an individual-level posterior for a particular non-experimental respondent, to predict her insurance-purchase decision or to evaluate the welfare consequences of a policy for someone with her observed characteristics, post-stratification has nothing to say, because it operates only on the marginal distribution and discards the joint information linking covariates to parameters. Conditional exchangeability does the job that post-stratification cannot, precisely because it places the demographic structure inside the model rather than alongside it.

References

Allenby, Greg M., and Peter E. Rossi. 1999. Marketing Models of Consumer Heterogeneity. *Journal of Econometrics* 89: 57–78.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. Lost in State Space: Are Preferences Stable? *International Economic Review* 49: 1091–112. [CrossRef]

Athey, Susan, and Guido W. Imbens. 2017. The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives* 31: 3–32. [CrossRef]

Blackburn, McKinley, Glenn W. Harrison, and E. Elisabet Rutström. 1994. Statistical Bias Functions and Informative Hypothetical Surveys. *American Journal of Agricultural Economics* 76: 1084–88. [CrossRef] [PubMed]

Bland, James R., Glenn W. Harrison, and Brian Monroe. 2026. *"The Hunt for Loss Aversion: A Bayesian Detective Story," Working Paper, Center for the Economic Analysis of Risk*. Atlanta: Robinson College of Business, Georgia State University.

Brooks, Stephen P., and Andrew Gelman. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7: 434–55. [CrossRef]

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76: 1–32. [CrossRef] [PubMed]

de Finetti, Bruno. 1937. La Prévision: Ses Lois Logiques, Ses Sources Subjectives. In *Annales de l'Institut Henri Poincaré*. English translation in Kyburg, Henry E., and Howard E. Smokler, eds. 1964. *Studies in Subjective Probability*. New York: Wiley, vol. 7, pp. 1–68.

Diaconis, Persi, and David Freedman. 1980. Finite Exchangeable Sequences. *Annals of Probability* 8: 745–64. [CrossRef]

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association* 9: 522–50. [CrossRef]

Gao, Xiaoxue Sherry, Glenn W. Harrison, and Rusty Tchernis. 2023. Behavioral welfare economics and risk preferences: A Bayesian approach. *Experimental Economics* 26: 273–303. [CrossRef]

Gelman, Andrew, and Donald B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7: 457–72. [CrossRef]

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Rubin Donald. 2013. *Bayesian Data Analysis*, 3rd ed. Boca Raton: Chapman & Hall/CRC.

Gronau, Quentin F., Alexandra Sarafoglou, Dora Matzke, Alexander Ly, Udo Boehm, Maarten Marsman, David S. Leslie, Jonathan J. Forster, Eric-Jan Wagenmakers, and Helen Steingroever. 2017. A Tutorial on Bridge Sampling. *Journal of Mathematical Psychology* 81: 80–97. [CrossRef] [PubMed]

Harrison, Glenn, Jimmy Martínez-Correa, Karlijn Morsink, Jia Min Ng, and J. Todd Swarthout. 2025. Welfare Consequences of the Compound Risks of Index Insurance. *Journal of Risk & Insurance*, 1–31. [CrossRef]

Harrison, Glenn W. 2025. *Causal Inferences Over Unobservables*. CEAR Working Paper 2025–02. Atlanta: Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

Harrison, Glenn W., and E. Elisabet Rutström. 2009. Expected Utility and Prospect Theory: One Wedding and a Decent Funeral. *Experimental Economics* 12: 133–58. [CrossRef]

Harrison, Glenn W., and Jia Min Ng. 2016. Evaluating the Expected Welfare Gain From Insurance. *Journal of Risk and Insurance* 83: 91–120.

Harrison, Glenn W., and J. Todd Swarthout. 2023. Cumulative Prospect Theory in the Laboratory: A Reconsideration. In *Models of Risk Preferences: Descriptive and Normative Challenges*. Edited by Glenn W. Harrison and Don Ross. Bingley: Emerald, Research in Experimental Economics, vol. 22, pp. 107–92.

Harrison, Glenn W., Jimmy Martínez-Correa, and J. Todd Swarthout. 2015. Reduction of Compound Lotteries with Objective Probabilities: Theory and Evidence. *Journal of Economic Behavior & Organization* 119: 32–55. [CrossRef]

Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutström. 2007. Estimating Risk Attitudes in Denmark: A Field Experiment. *Scandinavian Journal of Economics* 109: 341–68. [CrossRef]

Hernán, Miguel A., and James M. Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Hoffman, Matthew D., and Andrew Gelman. 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15: 1593–623.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

Johnson, Alicia A., Miles Q. Ott, and Dogucu Mine. 2012. *Bayes Rules! An Introduction to Applied Bayesian Modeling*. Boca Raton: CRC Press.

Kruschke, John K. 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd ed. Cambridge: Academic Press.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Lindley, Dennis V., and Adrian F. M. Smith. 1972. Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society, Series B* 34: 1–41. [CrossRef]

Loomes, Graham, and Robert Sugden. 1998. Testing Different Stochastic Specifications of Risky Choice. *Economica* 65: 581–98. [CrossRef]

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, 2nd ed. Boca Raton: CRC Press.

Nilsson, Håkan, Jörg Rieskamp, and Eric-Jan Wagenmakers. 2011. Hierarchical Bayesian Parameter Estimation for Cumulative Prospect Theory. *Journal of Mathematical Psychology* 55: 84–93. [CrossRef]

Phan, Du, Neeraj Pradhan, and Martin Jankowiak. 2019. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. *arXiv* arXiv:1912.11554.

Prelec, Drazen. 1998. The Probability Weighting Function. *Econometrica* 66: 497–527. [CrossRef]

Rossi, Peter E., and Greg M. Allenby. 1993. A Bayesian Approach to Estimating Household Parameters. *Journal of Marketing Research* 30: 171–82. [CrossRef]

Schwarz, Gideon. 1978. Estimating the Dimension of a Model. *Annals of Statistics* 6: 461–64. [CrossRef]

Vanpaemel, Wolf. 2010. Prior Sensitivity in Theory Testing: An Apologia for the Bayes Factor. *Journal of Mathematical Psychology* 54: 491–98. [CrossRef]

Vats, Dootika, and Christina Knudson. 2021. Revisiting the Gelman–Rubin diagnostic. *Statistical Science* 36: 518–29. [CrossRef]

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC. *Statistics and Computing* 27: 1413–32. [CrossRef]

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Burkner. 2021. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis* 16: 667–718. [CrossRef]

Wilcox, Nathaniel T. 2011. 'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice Under Risk. *Journal of Econometrics* 162: 89–104. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.