UNIVERSITY OF WARWICK

Economics

What Does Structural Noise Measure? Revealed-Preference Inconsistency and Model Misspecification

Ali Moghaddasi Kelishomi and Daniel Sgroi

August 2026
No: 1629

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Warwick Economics Research Papers

What Does Structural Noise Measure? Revealed-Preference Inconsistency and Model Misspecification*

Ali Moghaddasi Kelishomi[1] and Daniel Sgroi[2]

[1]Loughborough University and IZA

[2]University of Warwick and IZA

Abstract

Empirical comparisons of preferences often treat the dispersion parameter in a random utility model as noise around a correctly specified utility function. We show that this interpretation can fail: dispersion absorbs systematic behaviour that the fitted preference family cannot span. Using portfolio choices from four studies, we estimate disappointment-aversion random utility models and compare the dispersion parameter with revealed-preference diagnostics. A money-metric decomposition separates axiomatic inconsistency from the additional loss due to imposing DA-CRRA preferences. Both components independently predict dispersion, and model-relative misspecification remains strongly predictive among subjects who satisfy revealed-preference consistency. In our sharpest test, coefficients estimated on three studies predict dispersion in a held-out representative sample. A calibrated simulation illustrates the distinction. The dispersion parameter is therefore not a model-free measure of decision quality. It is jointly informative about choice inconsistency and preference-model fit, with implications for comparisons of recovered preferences and stochastic precision across people and environments.

Keywords: Revealed Preference, Random Utility Models, Model Misspecification, Structural Noise, Choice Consistency.

JEL classification: D81, D11, D12, C91.

*We are very grateful for helpful discussions with Matthew Polisson, Hong Il Yoo and David Gill. Ali Moghaddasi Kelishomi, Loughborough Business School, Loughborough University, Loughborough, United Kingdom. a.moghaddasi@lboro.ac.uk. Daniel Sgroi, Department of Economics, University of Warwick, Coventry, United Kingdom. daniel.sgroi@warwick.ac.uk.

1 Introduction

Structural estimates of preferences are routinely compared across people, treatments, and populations. Those comparisons rely on a consequential division of the data: a deterministic utility function is interpreted as preferences, and the residual dispersion in choice is interpreted as imperfect implementation around those preferences. If the utility family is correctly specified, that division is natural. If it is not, systematic choice patterns that the model cannot represent are also assigned to the dispersion term. Apparent differences in "choice noise" may then reflect differences in model fit, and preference comparisons may be most fragile precisely where the fitted model reports high dispersion.

This paper asks what the estimated dispersion parameter in a structural random utility model actually measures. The answer matters beyond terminology. Dispersion parameters are used to characterise decision quality, compare the precision of behaviour across tasks, and accommodate deviations from deterministic maximisation while recovering preference parameters. Those interpretations require a distinction between two sources of residual variation: choices that are internally inconsistent and choices that are coherent but not spanned by the imposed preference family. A single stochastic scale parameter can absorb both.

We distinguish these sources empirically by bringing two measurement traditions to the same choices. Revealed-preference methods assess whether choices can be rationalised without imposing a parametric utility family. Structural random utility models instead impose a deterministic preference family and represent departures from its optimum probabilistically. Portfolio-allocation experiments with intersecting linear budgets are unusually well suited to comparing the two: the same rich choice sets support axiomatic consistency tests, money-metric measures of parametric fit, and individual-level structural estimation.

We combine four such studies (Choi et al., 2007, 2014; Halevy et al., 2018; Moghaddasi Kelishomi and Sgroi, 2024). For each subject, we estimate a disappointment-aversion random utility model (DA–RUM), based on Gul (1991), on finite menus constructed using the GRID method of Polisson et al. (2020). The model recovers CRRA curvature, disappointment aversion, and a dispersion parameter $\hat{\sigma}$. A menu-specific normalisation, motivated by Wilcox (2011), expresses dispersion relative to the local utility spread and substantially reduces the mechanical dependence of $\hat{\sigma}$ on cardinal utility scale.

The analysis proceeds in three steps. First, we establish that $\hat{\sigma}$ has the expected relationship with axiomatic inconsistency. GARP, fGARP, and EU consistency are all strongly negatively associated with structural dispersion. A split-sample exercise estimates $\hat{\sigma}$ on one half of a subject's choices and GARP on the other; the relationship remains negative, and the odd–even correlation of $\hat{\sigma}$ is 0.749. The dispersion estimate is therefore not merely reproducing violations in the identical observations used to fit it.

Second, and centrally, axiomatic inconsistency does not exhaust what $\hat{\sigma}$ captures. We use the Money Metric Index of Halevy et al. (2018), which decomposes the loss from imposing DA-CRRA into Varian inefficiency—the cost of internal revealed-preference inconsistency—and an additional DA-CRRA misspecification residual. Both components independently predict $\hat{\sigma}$. The misspecification component remains large and significant among subjects with no detected revealed-preference inconsistency. Coherent choices can therefore receive high structural dispersion when they are poorly described by the fitted preference family.

Third, the mapping based on this decomposition transports across samples. We estimate the mapping from the two money-metric components to $\hat{\sigma}$ in Studies 1, 3, and 4, and apply it without re-estimating the prediction equation to the representative CentERpanel sample in Study 2. The strict held-out prediction has an out-of-sample R^2 of 0.816. A model using Varian inefficiency alone has an out-of-sample R^2 of 0.330 on the same test observations. Thus, model-relative lack of fit is not a semantic relabelling of inconsistency; it accounts for substantial, stable variation in the structural residual.

A calibrated simulation isolates the mechanism. Under the correctly specified DA data-generating process, lower revealed-preference consistency raises $\hat{\sigma}$. Holding consistency fixed, a structured heuristic outside DA-CRRA raises $\hat{\sigma}$ sharply, whereas expected utility, which is nested in DA, does not produce a comparable increase. Cognitive ability provides a complementary external check: it predicts lower $\hat{\sigma}$, but not the recovered preference parameters, and most of this association overlaps with GARP consistency.

The contribution is an empirical discipline for interpreting structural residuals, not a new estimator or a model-free decomposition of $\hat{\sigma}$. The estimates show that a standard dispersion parameter is jointly informative about choice consistency and the adequacy of the fitted preference family. Consequently, $\hat{\sigma}$ is neither a pure measure of decision quality nor a direct measure of estimation uncertainty in the recovered preference parameters. Its meaning is model-relative. This distinction matters whenever researchers compare estimated preferences or stochastic precision across groups: a change in dispersion can arise because behaviour became less consistent, because the chosen utility family fits one group worse, or both. Preference parameters and $\hat{\sigma}$ should therefore be reported together, but diagnosing the source of dispersion requires an external fit benchmark such as the one used here.

Related Literature

This paper contributes to three strands of research.

The first is the revealed-preference literature on axiomatic consistency in experimental choice. Afriat (1967) establishes the foundational equivalence between GARP and utility maximisation, and Varian (1982) develops the nonparametric testing framework.

Choi et al. (2007) introduce the portfolio-choice experiment used throughout this paper and document substantial heterogeneity in CCEI scores. Choi et al. (2014) extend the analysis to a large representative sample and show that consistency is positively associated with wealth and education. Polisson et al. (2020) develop the GRID discretisation method that underlies both the revealed-preference analysis and the structural estimation here. The revealed-preference literature has also produced a range of scalar indices that embody different notions of distance from rationality: the CCEI and Varian Inefficiency Index measure the budget contraction or adjustment required to restore consistency; the swaps index of Apesteguia and Ballester (2015) counts the swaps required to reconcile observed choices with a candidate preference ranking; subset-based measures such as the Houtman–Maks index identify the largest rationalisable subset; and Echenique et al. (2023) measure the minimal perturbation to marginal conditions required for expected-utility rationalisation. We use CCEI-type indices for GARP, fGARP, and EU, together with the Varian index, treating them as complementary diagnostics of different dimensions of departure from optimal choice rather than equivalent representations of a single underlying rationality score.

The second strand is the structural estimation literature. Standard approaches combine parametric utility functions with stochastic choice rules to recover preference parameters from experimental data (McFadden, 1972; Andersen et al., 2008; Wilcox, 2011). The Fechner-logit specification, in which the probability of choosing an alternative increases in its utility advantage over competitors, is widely used in this literature. Wilcox (2011) shows that the raw utility scale can affect the relationship between deterministic risk aversion and stochastic choice probabilities in undesirable ways, and proposes contextual utility normalisation as a correction. Our menu-specific analogue of this procedure is described in Section 2. More recent work has developed identification results for preference parameters in richer experimental environments (e.g., Lau and Yoo, 2025). Structural estimation has focused predominantly on recovering preference parameters, while the interpretation of the estimated dispersion parameter has received comparatively little systematic attention.

A closely related paper is Halevy et al. (2018), who develop a deterministic preference recovery framework grounded in revealed-preference theory. Their central object, the Money Metric Index, decomposes into the Varian Inefficiency Index, capturing irreducible revealed-preference inconsistency, and a parametric misspecification residual, capturing the additional money-metric cost of restricting attention to a particular functional family. We use their decomposition as a diagnostic tool by relating the ML-estimated $\hat{\sigma}$ to the Varian inefficiency and misspecification components of the MMI. The two estimators employ different objective functions: Halevy et al. minimise a money-metric budget-adjustment criterion, whereas we maximise a Fechner–logit likelihood. Because the two procedures use different objective functions, the relationship between the MMI

components and $\hat{\sigma}$ is empirical rather than algebraic.

The third strand is the literature on stochastic choice and measurement error in preference elicitation. Apesteguia and Ballester (2018) study whether stochastic choice probabilities vary monotonically with the underlying preference parameter. They show that standard random-utility specifications used in risk and time preference estimation can violate this property, potentially creating identification and estimation problems, whereas random-parameter models satisfy it. Our concern is distinct but related: within a Fechner–logit specification, the cardinal scale of utility affects the mapping from preference parameters to choice probabilities. Following Wilcox (2011), we therefore normalise utility within each menu to reduce the extent to which estimated dispersion reflects arbitrary utility-scale variation.

Andersson et al. (2020) show that in binary lottery tasks cognitive ability predicts noisy behaviour rather than risk preferences, and that failing to model decision noise can generate a spurious correlation between cognitive ability and risk aversion. Our results replicate and extend this finding to a richer continuous-menu environment. More broadly, they contribute to a growing literature emphasising that cognitive limitations affect the precision with which individuals implement decisions rather than the content of underlying preferences (Enke, 2024). A comprehensive review of elicitation methods and their associated measurement challenges is provided by Chapman and Fisher (2025).

The remainder of the paper proceeds as follows. Section 2 presents the framework and empirical strategy. Section 3 describes the empirical setting, revealed-preference measures, GRID construction, samples, and descriptive estimates. Section 4 relates the DA–RUM estimates to axiomatic consistency. Section 5 examines whether structural dispersion also reflects DA-CRRA-relative misspecification after conditioning on Varian inefficiency. Section 6 presents the held-out cross-study prediction exercise. Section 7 reports validation and robustness evidence, including convergence analysis, the calibrated simulation, and the relationship with cognitive ability. Section 8 discusses the implications for comparisons of recovered risk attitudes. Section 9 concludes.

2 Framework and Empirical Strategy

The fitted structural model maps each subject's choices into preference parameters and residual dispersion. The empirical strategy then uses revealed-preference objects, constructed under weaker preference restrictions, to characterise that residual. This separation is important: the paper does not claim that the random utility model identifies inconsistency and misspecification as distinct latent shocks. Instead, it asks whether external diagnostics for those two sources explain distinct variation in the single estimated dispersion parameter.

2.1 Framework

The portfolio task generates choices from continuous budget sets. We evaluate both the revealed-preference diagnostics and the multinomial likelihood on discretised feasibility sets constructed using the GRID method of Polisson et al. (2020). The construction preserves the relevant rationalisability restrictions while providing finite menus for structural estimation. Section 3 describes the menus in detail.

2.2 Structural Model and Utility-Scale Normalisation

We model preferences using Gul (1991)'s symmetric disappointment-aversion specification, which permits a kink at certainty. For each grid point $g = (x_1, x_2) \in G_{it}$, utility is defined as

$$U(g; \rho, a) = \min \{au(x_1; \rho) + u(x_2; \rho), u(x_1; \rho) + au(x_2; \rho)\}, \qquad a \geq 1, \tag{1}$$

where

$$u(x; \rho) = \begin{cases} \dfrac{x^{1-\rho}}{1-\rho}, & \rho \neq 1, \\ \log x, & \rho = 1, \end{cases} \tag{2}$$

is the CRRA utility index.

To account for stochastic choice behaviour, we employ a Fechner–logit specification with a menu-specific normalisation of the utility index. For individual i in decision round t, define the within-menu scale

$$S_{it}(\rho, a) = \mathrm{sd}\left\{U(h; \rho, a) - \bar{U}_{it}(\rho, a) : h \in G_{it}\right\}, \tag{3}$$

where $\bar{U}_{it}(\rho, a)$ denotes mean utility across all grid points in G_{it}. The menu-specific normalisation is important because the geometry and density of the discretised menus vary across individuals and decisions. Dividing utility differences by their within-menu standard deviation expresses σ relative to the local utility spread in each menu. This reduces the extent to which differences in $\hat{\sigma}$ reflect variation in the cardinal scale of utility rather than stochastic choice dispersion.

This adjustment addresses a general scale problem in stochastic choice models. Choice probabilities depend on cardinal utility differences, so changes in preference curvature can alter the estimated dispersion even when the underlying degree of stochastic choice is unchanged. Wilcox (2011) proposes contextual normalisation for binary lottery choices. Our within-menu standardisation provides an analogue for discretised linear-budget menus: each alternative's utility is expressed relative to the utility spread of the menu in which it appears. This reduces the extent to which $\hat{\sigma}$ mechanically reflects the raw cardinal scale induced by (ρ, a).

The probability of selecting grid point g from menu G_{it} is

$$P_{it}(g \mid G_{it}) = \frac{\exp\left[\dfrac{U(g;\rho,a) - \bar{U}_{it}(\rho,a)}{S_{it}(\rho,a)\sigma}\right]}{\displaystyle\sum_{h \in G_{it}} \exp\left[\dfrac{U(h;\rho,a) - \bar{U}_{it}(\rho,a)}{S_{it}(\rho,a)\sigma}\right]}. \tag{4}$$

The normalised utility difference entering the choice probability is dimensionless. Appendix A shows that the menu-specific adjustment reduces curvature-induced variation in the implied dispersion parameter by approximately 75 percent across representative budgets. The normalisation therefore substantially reduces the mechanical utility-scale channel and improves the comparability of $\hat{\sigma}$ across menus and individuals, while preserving the joint estimation of preference and dispersion parameters.

Here, σ is the structural dispersion parameter: lower values imply that choice probabilities are more concentrated around the utility-maximising grid point, while higher values imply greater dispersion across alternatives. Under the menu-specific normalisation, dispersion is measured relative to the local utility scale of each menu. Choice precision is represented by $1/\sigma$, not by σ itself.

2.3 Parameter Recovery and the Money-Metric Decomposition

Parameters (ρ, a, σ) are estimated separately for each individual by maximum likelihood. Let g_{it} denote the chosen grid point from menu G_{it} in round t. The individual log-likelihood is

$$L_i(\rho, a, \sigma) = \sum_{t=1}^{T_i} \log \left[\frac{\exp\left(\dfrac{U(g_{it};\rho,a) - \bar{U}_{it}(\rho,a)}{S_{it}(\rho,a)\sigma}\right)}{\displaystyle\sum_{h \in G_{it}} \exp\left(\dfrac{U(h;\rho,a) - \bar{U}_{it}(\rho,a)}{S_{it}(\rho,a)\sigma}\right)} \right]. \tag{5}$$

Estimation therefore jointly recovers risk curvature ρ, disappointment aversion a, and structural dispersion σ for each individual.

To interpret the recovered dispersion parameter, we combine these likelihood estimates with the money-metric decomposition of Halevy et al. (2018), described in Section 3.3. For the DA-CRRA family, the decomposition is

$$\text{MMI}_{DA\text{-}CRRA} = \text{Varian Inefficiency} + \text{DA-CRRA Misspecification}. \tag{6}$$

The first component captures internal revealed-preference inconsistency. The second is the additional money-metric loss generated by restricting rationalisation to the DA-CRRA family.

The two procedures use the same choices but different recovery criteria. Halevy et al. recover approximate preferences by minimising money-metric ranking incompatibility, whereas the DA-RUM recovers preferences and dispersion jointly by maximising a stochastic-choice likelihood. We do not use the MMI to estimate (ρ, a, σ). We use its two components as deterministic diagnostics for examining whether the likelihood-based $\hat{\sigma}$ is related only to revealed-preference inconsistency or also to DA-CRRA-relative lack of fit. Because the objective functions differ, the relationship between these objects is empirical rather than algebraic.

2.4 Empirical Mapping Strategy

The empirical analysis proceeds as follows. Section 4 establishes the initial mapping between $\hat{\sigma}$ and GARP, fGARP, and EU consistency. Section 5 then introduces the two MMI components jointly and examines whether DA–CRRA-relative misspecification remains predictive after conditioning on Varian inefficiency, including among subjects with no detected revealed-preference inconsistency. Section 6 estimates the same mapping using Studies 1, 3, and 4 and evaluates its ability to predict $\hat{\sigma}$ in held-out Study 2 relative to a Varian-only benchmark. Section 7 examines estimation robustness and provides supplementary validation through a calibrated model-spanning simulation and the relationship between cognitive ability and the recovered parameters. Section 8 then examines how structural dispersion varies across the distribution of recovered risk attitudes.

3 Empirical Setting, Measures, and Samples

3.1 The Choi-Type Budgetary Choice Experiment

Figure 1a illustrates the Choi-type portfolio task introduced by Choi et al. (2007, 2014). In each decision round t, subjects allocate wealth between two Arrow securities that pay in two equiprobable states of the world. A portfolio is represented by its state-contingent payoffs (x_1, x_2) and is chosen subject to

$$p_{1t}x_1 + p_{2t}x_2 = 1, \qquad x_1, x_2 \geq 0,$$

where the state prices p_{1t} and p_{2t}, and hence their relative price, vary across rounds.

The graphical interface allows subjects to select any allocation on the budget frontier. Each observation therefore corresponds to a choice from a continuum of alternatives rather than from a discrete menu. Repeating the task across different price vectors generates choices from intersecting budget sets, as depicted in Figure 1a.

A key feature of the design is that the chosen bundle is revealed preferred to every affordable bundle in the corresponding budget set. This ranking information forms the

basis for both the revealed-preference consistency tests and the structural estimation developed below.

3.2 Revealed-Preference Efficiency Indices

Departures from rational choice are quantified using revealed-preference efficiency indices that measure how far observed behaviour is from exact utility maximisation.

GARP and the Afriat CCEI. Consistency with utility maximisation is equivalent to satisfaction of the Generalized Axiom of Revealed Preference (GARP). Afriat's Critical Cost Efficiency Index (CCEI) is the largest scalar $e \in (0, 1]$ such that the data satisfy GARP when all budgets are scaled by e. Equivalently, $1 - e$ is the smallest uniform proportional contraction required to restore GARP. A value of $e = 1$ indicates exact consistency, while lower values indicate more severe violations.

fGARP and EU CCEI. In risky-choice settings, additional preference restrictions can be imposed. fGARP requires choices to be rationalisable by continuous preferences that are monotone with respect to first-order stochastic dominance. EU imposes the further restriction of expected-utility maximisation. Following Polisson et al. (2020), we compute CCEI-type scores for these restrictions using the GRID method.

The Varian index. Varian (1990) generalises the Afriat measure by allowing the adjustment factor to vary across budgets rather than imposing a common scalar. The Varian index aggregates the smallest budget-specific adjustments required to restore GARP. It therefore permits the degree of inefficiency to vary across choice situations. Throughout the analysis, both the Varian index and the MMI use the normalised root-mean-square (SSQ) aggregator employed in the main application of Halevy et al. (2018).

3.3 Money-Metric Inconsistency and Misspecification

Halevy et al. (2018) develop a preference-recovery framework grounded in revealed-preference theory. For a candidate utility function, their Money Metric Index (MMI) aggregates the smallest budget-specific adjustments required to remove incompatibilities between the ranking revealed by observed choices and the ranking induced by that utility function.

MMI and preference recovery. For a given utility function $u(\cdot)$, the MMI adjusts each budget until no affordable alternative is ranked strictly above the chosen bundle. Minimising this loss over a utility family recovers the member of that family that most closely represents the revealed ranking under the money-metric criterion. Halevy et al.

apply this procedure to symmetric disappointment aversion with a CRRA utility index and compare it with a distance-based recovery criterion in commodity space.

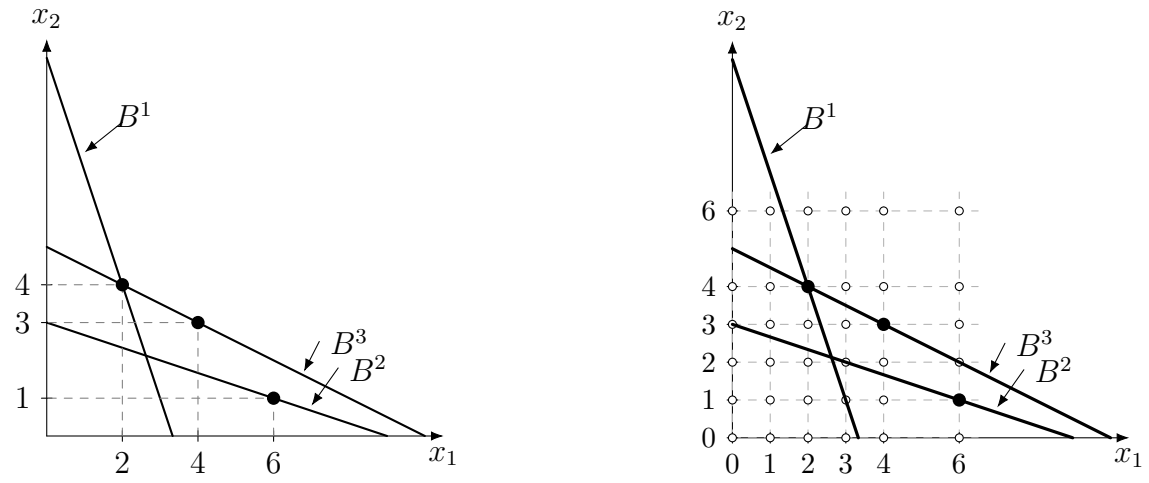
Inconsistency and misspecification. Using the same aggregator for both components, Halevy et al. show that the minimum money-metric loss for the DA-CRRA family can be decomposed as

$$\text{MMI}_{DA\text{-}CRRA} = \text{Varian Inefficiency} + \text{DA-CRRA Misspecification}.$$

The Varian component measures internal revealed-preference inconsistency. The residual measures the additional money-metric loss generated by restricting rationalisation to the DA-CRRA family rather than allowing the broader class of admissible utility functions. This decomposition supplies the deterministic benchmark for the paper's central test.

3.4 Finite-Menu Construction and Structural Menus

Figure 1 illustrates how choices across intersecting linear budgets generate revealed-preference comparisons. The structural likelihood requires a finite set of alternatives for each decision, whereas the experimental interface permits any allocation on the continuous budget line. We obtain finite menus using the Generalized Restriction of Infinite Domains (GRID) method of Polisson et al. (2020).



(a) Dataset with three observations (b) Dataset with grid inserted

Figure 1: Illustration of a Choi-type portfolio task with multiple budgets and the grid method. *Source: Adapted from Polisson et al. (2020).*

The GRID method begins by constructing a set $\mathcal{X}$ which consists of zero and every consumption level appearing in the subject's observed choices:

For each individual i, define

$$\mathcal{X}_i = \{0\} \cup \{x_{ist} : t = 1, \ldots, T_i,\ s \in \{1, 2\}\}, \qquad \mathcal{G}_i = \mathcal{X}_i^2. \tag{7}$$

For each decision t, the structural menu is the set of undominated grid points contained in the budget set. The resulting menu preserves the relevant rationalisability restrictions while making individual-level multinomial likelihood estimation computationally feasible. As shown in Polisson et al. (2020), a dataset is rationalisable by a broad class of models (including EU and DA) if and only if it is rationalisable on the finite grid $G = \mathcal{X}^2$ generated by the Cartesian product of these observed levels. Appendix B verifies that the main results are qualitatively unchanged when the endogenous GRID menus in Studies 3 and 4 are replaced with exogenous uniform grids.

Figure 1b illustrates this discretization process. The continuous budget line is overlaid with the grid G, represented by the intersection of the horizontal and vertical lines corresponding to the elements of $\mathcal{X}$. The finite choice menu for each round t is then defined as the intersection of the grid and the budget set, $G_{it} = B_{it} \cap G_i$ where

$$B_{it} = \left\{x \in \mathbb{R}^2_+ : p_{1t}x_1 + p_{2t}x_2 \leq 1\right\}.$$

By selecting the undominated bundles within this intersection, we create a discrete set of alternatives that preserves the relevant revealed-preference information of the original continuous problem. [1]

The CRRA index is singular at zero for parts of the admissible parameter space. We therefore replace zero consumption at boundary grid points with a small positive floor, ε, when evaluating utility. This adjustment is applied only inside the likelihood calculation; the observed allocation and budget remain unchanged. The value of $\varepsilon = 10^{-6}$ is fixed across subjects, studies, and parameter evaluations.

We treat G_{it} as the multinomial choice menu. The same finite-menu foundation therefore underlies the revealed-preference diagnostics and structural likelihood. Appendix B replaces the choice-dependent GRID menus in Studies 3 and 4 with exogenous uniform grids and preserves the principal relationships.

3.5 Samples

The four studies use the same graphical portfolio-allocation task but differ sharply in subject composition and decisions per subject. This common task limits design hetero-

[1] For individual i in round t, a grid point $\mathbf{x} \in B_{it} \cap \mathcal{G}_i$ is statewise dominated if there exists $\mathbf{y} \in B_{it} \cap \mathcal{G}_i$ such that $y_s \geq x_s$ for every state s, with strict inequality for at least one state. Because DA-CRRA utility is strictly increasing in each state payoff, a dominated point cannot maximise utility whenever its dominating point is feasible. Since the grid is finite, maximisation over its undominated points is therefore equivalent to maximisation over the full grid. We define G_{it} as this undominated subset and evaluate the multinomial likelihood in Equation (5) on the resulting reduced menu.

geneity, while the student, representative-population, and field samples provide useful variation in data density and revealed-preference consistency.

Study 1. Study 1 uses the symmetric treatment of Choi et al. (2007). The original study recruited 93 undergraduate students and staff at the University of California, Berkeley. Our sample comprises the 47 subjects in the treatment in which the two states occurred with equal probability. Each subject completed 50 decision problems.

Study 2. Study 2 uses the 1,182 CentERpanel participants studied by Choi et al. (2014), a representative sample of the Dutch-speaking population of the Netherlands. Each subject completed 25 decisions using an online version of the Study 1 interface.

Study 3. Study 3 uses the 203 primarily undergraduate participants studied by Halevy et al. (2018) at the University of British Columbia. We use the first part of the experiment, in which each subject completed 22 portfolio-allocation decisions across equiprobable states. The budget sets emphasised moderate price ratios to improve identification of first-order risk attitudes.

Study 4. Study 4 uses the 218 participants in the Iranian kidney market studied by Moghaddasi Kelishomi and Sgroi (2024). The sample consists of living unrelated kidney donors, typically from the lower end of the income distribution. Each participant completed 25 decisions using the same graphical portfolio-allocation task. The study therefore provides a field sample drawn from an institutional setting that differs sharply from the student and representative-panel samples.

Table 1: Summary of Choi-Type Datasets Used in the Analysis

Study	Reference	Subject Pool	N	# Tasks
Study 1	Choi et al. (2007)	UC Berkeley (Students/Staff)	47	50
Study 2	Choi et al. (2014)	CentERpanel (Dutch Representative)	1,182	25
Study 3	Halevy et al. (2018)	UBC (Students)	203	22
Study 4	Moghaddasi Kelishomi and Sgroi (2024)	Iranian Kidney Donors	218	25

Notes. Each sample offers distinct advantages. Study 1 is characterised by a high task density (50 rounds). Study 2 provides the largest sample size and the most representative demographic variation, although with a lower number of tasks per subject (25 rounds). Study 3 uses a shorter sequence of 22 rounds and specifically emphasises moderate price ratios to better identify the role of first-order risk aversion. Study 4 offers a unique sample consisting of actual donors in the Iranian kidney market.

Table 1 summarises the samples. Analysis counts vary with convergence, parameter trimming, and availability of the money-metric measures; every results table states its own restrictions.

3.6 Descriptive Structural Estimates

We first describe the individual-level DA–RUM estimates. The reported distributions retain converged estimates with $\hat{a} < 15$, excluding a region in which disappointment aversion is weakly identified.

Table 2: Non-Parametric Consistency and Recovered Structural Parameters

Study	GARP	fGARP	EU	$\hat{\rho}$	$\hat{a}$	$\hat{\sigma}$
Study 1 (Choi 07)	0.916	0.879	0.872	0.266	1.734	0.117
	(0.104)	(0.129)	(0.125)	(0.268)	(0.790)	(0.105)
Study 2 (Choi 14)	0.884	0.754	0.752	0.473	2.179	0.200
	(0.132)	(0.210)	(0.208)	(0.351)	(2.110)	(0.192)
Study 3 (Halevy 18)	0.975	0.948	0.945	0.346	1.710	0.096
	(0.056)	(0.076)	(0.075)	(0.317)	(1.386)	(0.083)
Study 4 (MS 24)	0.881	0.742	0.741	0.500	1.292	0.215
	(0.117)	(0.183)	(0.183)	(0.304)	(0.367)	(0.174)

Notes. The table reports means, with standard deviations in parentheses, for subjects with converged DA–RUM estimates and $\hat{a} < 15$. The revealed-preference indices for Studies 1-4 are computed using the replication code of Polisson et al. (2020).

Table 2 shows the first descriptive pattern. Study 3 has the highest mean consistency and the lowest mean $\hat{\sigma}$; Studies 2 and 4 have lower consistency and higher dispersion. Because the studies also differ in subject composition, task density, and budget environments, the cross-study means alone cannot attribute this ordering to differences in consistency. The individual-level regressions therefore include study fixed effects. The preference parameters vary across studies as well, but not in the same order as dispersion.

Table 3 reports the full distributions. Median $\hat{\rho}$ ranges from 0.214 in Study 1 to 0.458 in Study 4, while the distribution of disappointment aversion also varies substantially across studies. Structural dispersion is strongly right-skewed: median $\hat{\sigma}$ ranges from 0.076 in Study 1 and 0.081 in Study 3 to 0.165 in Study 2 and 0.188 in Study 4. These descriptive differences motivate, but cannot establish, the paper's interpretation. The following sections use within-study variation and the money-metric decomposition to examine whether dispersion is associated with inconsistency, DA-CRRA-relative lack of fit, or both.

Table 3: Distribution of Recovered DA–RUM Parameters across Studies

Study	Var	Mean	SD	p10	p25	p50	p75	p90	p95	N
Study 1	$\hat{\rho}$	0.266	0.268	0.021	0.098	0.214	0.364	0.481	0.522	24
	$\hat{a}$	1.734	0.790	1.092	1.361	1.468	1.720	3.271	3.440	
	$\hat{\sigma}$	0.117	0.105	0.027	0.044	0.076	0.160	0.312	0.317	
Study 2	$\hat{\rho}$	0.473	0.351	0.000	0.236	0.429	0.655	0.937	1.092	644
	$\hat{a}$	2.179	2.110	1.000	1.000	1.294	2.227	4.770	6.097	
	$\hat{\sigma}$	0.200	0.192	0.004	0.048	0.165	0.303	0.431	0.507	
Study 3	$\hat{\rho}$	0.346	0.317	0.000	0.141	0.246	0.550	0.697	0.910	105
	$\hat{a}$	1.710	1.386	1.000	1.132	1.470	1.907	2.331	3.083	
	$\hat{\sigma}$	0.096	0.083	0.008	0.031	0.081	0.134	0.185	0.238	
Study 4	$\hat{\rho}$	0.500	0.304	0.157	0.265	0.458	0.658	0.885	0.984	133
	$\hat{a}$	1.292	0.367	1.000	1.000	1.177	1.459	1.725	1.955	
	$\hat{\sigma}$	0.215	0.174	0.021	0.068	0.188	0.313	0.443	0.596	

Notes. The table reports study-specific distributions of the individual-level DA–RUM estimates. The sample is restricted to subjects whose maximum-likelihood estimation converged and whose recovered disappointment-aversion parameter satisfies $\hat{a} < 15$. N denotes the number of subjects satisfying these restrictions in each study and is common to the $\hat{\rho}$, $\hat{a}$, and $\hat{\sigma}$ rows. SD denotes the standard deviation, and $p10$, $p25$, $p50$, $p75$, $p90$, and $p95$ denote the corresponding percentiles.

4 Does Structural Dispersion Measure Axiomatic Inconsistency?

The first test asks whether $\hat{\sigma}$ is associated with the residual it is most naturally expected to reflect: internally inconsistent choice. Table 4 relates the GARP CCEI to the recovered preference parameters in Columns (1) and (2), and relates $\hat{\sigma}$ separately to the GARP, fGARP, and EU CCEI scores in Columns (3)–(5). Column (6) allows the GARP–dispersion relationship to vary across studies.

All three consistency measures are strongly negatively associated with dispersion. In Column (3), a 0.10 increase in the GARP CCEI is associated with a 0.104 reduction in $\hat{\sigma}$. The coefficients on the fGARP and EU CCEI scores are also negative and precisely estimated. In Column (6), the implied GARP slope remains negative in every study, although its magnitude varies. Differences in budget geometry, task density, sample composition, and the distribution of consistency may all contribute to this heterogeneity. The robust conclusion is narrower: choices that are closer to axiomatic consistency receive less residual dispersion from the DA-RUM.

Because both objects summarise the same choices, these regressions establish an empirical mapping rather than a causal relationship. Appendix Table A4 removes exact

observational overlap in Study 1: we estimate $\hat{\sigma}$ on odd rounds and the GARP CCEI on even rounds, and then reverse the split. Complementary-split GARP remains negatively associated with $\hat{\sigma}$, while the odd–even correlation of $\hat{\sigma}$ is 0.749. Although the split samples are small, the result shows that the baseline relationship is not driven solely by using identical rounds to construct both objects. Having established that dispersion tracks axiomatic inconsistency, the next section asks whether inconsistency exhausts its empirical content.

Table 4: DA-RUM Parameters and Revealed-Preference Consistency

	(1) $\hat{\rho}_i$	(2) $\hat{a}_i$	(3) $\hat{\sigma}_i$	(4) $\hat{\sigma}_i$	(5) $\hat{\sigma}_i$	(6) $\hat{\sigma}_i$
GARP	0.814***	3.238***	-1.041***			-0.777***
	(0.076)	(0.457)	(0.040)			(0.101)
fGARP				-0.729***		
				(0.034)		
EU					-0.737***	
					(0.035)	
Study 2	0.233***	0.549**	0.049***	-0.009	-0.005	0.294***
	(0.052)	(0.218)	(0.015)	(0.016)	(0.016)	(0.102)
Study 3	0.032	-0.214	0.040**	0.028*	0.033*	0.311*
	(0.059)	(0.240)	(0.016)	(0.017)	(0.017)	(0.161)
Study 4	0.263***	-0.327	0.060***	-0.003	0.001	0.314***
	(0.056)	(0.203)	(0.018)	(0.018)	(0.018)	(0.120)
Study 2 × GARP						-0.268**
						(0.111)
Study 3 × GARP						-0.294*
						(0.167)
Study 4 × GARP						-0.277**
						(0.129)
Observations	906	906	906	906	906	906

Notes. GARP, fGARP, and EU denote the corresponding CCEI-type efficiency scores. All regressions are estimated on converged DA-RUM estimates with $\hat{a}_i < 15$. Robust standard errors are reported in parentheses. Study 1 is the omitted category. Each revealed-preference consistency measure is entered separately because the three scores are highly collinear. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

5 Structural Dispersion Also Captures Model Misspecification

The paper's central test separates internal inconsistency from coherent choice that DA-CRRA cannot span. The MMI provides that separation: its Varian component measures revealed-preference inconsistency, while the residual is the additional money-metric loss from imposing DA-CRRA. Table 5 enters both components as predictors of $\hat{\sigma}$.

Table 5: Structural Dispersion, Varian Inefficiency, and DA-CRRA Misspecification

	(1) Full Sample	(2) Refined Sample	(3) Refined Sample	(4) Consistent	(5) Refined Consistent
	$\hat{\sigma}_i$	$\hat{\sigma}_i$	$\hat{\sigma}_i$	$\hat{\sigma}_i$	$\hat{\sigma}_i$
Misspecification	1.205***	2.624***	2.282***	2.678***	2.678***
	(0.248)	(0.158)	(0.231)	(0.697)	(0.697)
Varian Inefficiency			1.877***		
			(0.231)		
Study 2	0.052**	0.000	-0.016***	0.005	0.005
	(0.024)	(0.008)	(0.005)	(0.012)	(0.012)
Study 3	-0.030	0.009	0.012**	0.021	0.021
	(0.024)	(0.008)	(0.006)	(0.013)	(0.013)
Study 4	0.050*	-0.010	-0.027***	-0.034	-0.034
	(0.029)	(0.011)	(0.008)	(0.021)	(0.021)
Observations	906	770	770	208	208

Notes. Misspecification equals MMI minus Varian Inefficiency. Column (3) enters both components jointly. The Refined Sample excludes weakly approximated Varian indices following Halevy et al. (2018). The Consistent Sample retains subjects who satisfy the revealed-preference criterion. Column (5) additionally applies the refined-sample restriction to the consistent-subject sample. Because no further observations are excluded by this restriction, Columns (4) and (5) contain the same 208 subjects and produce identical estimates. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The joint specification in Column (3) is the central result. Varian inefficiency and DA-CRRA misspecification are both positive and precisely estimated. Adding Varian inefficiency reduces the misspecification coefficient from 2.624 to 2.282, but leaves it highly significant. A 0.01 increase in the misspecification residual is associated with an increase of approximately 0.023 in $\hat{\sigma}$. Internal inconsistency is therefore important, but it does not exhaust the residual assigned by the structural model.

Columns (4) and (5) provide a sharper separation. Among 208 subjects with no detected Varian inefficiency, DA-CRRA misspecification remains large and significant.

These subjects pass the revealed-preference criterion, yet coherent choice patterns that are farther from DA-CRRA receive higher dispersion. Appendix G preserves the result under exact GARP and high fGARP and EU thresholds.

This result is deliberately model-relative. It does not establish an absolute quantity called "misspecification," nor does it identify the psychological source of residual choice. It shows that the stochastic scale of a DA-CRRA model systematically absorbs the additional fit loss created by restricting preferences to DA-CRRA. That is precisely why $\hat{\sigma}$ cannot be interpreted as a model-free measure of decision quality. The next section asks whether this mapping is stable enough to predict dispersion in a different sample.

6 Cross-Sample Prediction of Structural Dispersion

An in-sample association could reflect sample-specific covariance among three summaries of the same choices. We therefore ask whether the decomposition learned in Studies 1, 3, and 4 predicts $\hat{\sigma}$ in Study 2, the representative CentERpanel sample. No Study 2 observation enters estimation of the strict prediction equation.

The decomposition model regresses $\hat{\sigma}_i$ on Varian inefficiency, DA-CRRA misspecification, and indicators for Studies 3 and 4. A Varian-only benchmark uses the identical refined training and test samples but excludes misspecification. Their difference in performance isolates the incremental predictive content of model-relative fit.

For each model, Option A is the strict held-out forecast: Study 2 receives the omitted Study 1 intercept, and no Study 2 outcome is used to alter the prediction. Option B is a complementary calibration exercise that adds the mean Study 2 residual from Option A to every predicted value. This adjustment assesses calibration in levels after a single test-sample mean correction. Rankings, correlations, and cross-subject dispersion are unchanged by construction.

Table 6 reports strong held-out performance. The strict decomposition forecast has an out-of-sample R^2 of 0.816, a Pearson correlation of 0.913, and a Spearman correlation of 0.930. Re-centring raises R^2_{OOS} to 0.832 and removes the rejection of the Mincer–Zarnowitz calibration restriction. On the same test observations, the strict Varian-only model has an out-of-sample R^2 of 0.330. Adding misspecification therefore raises strict out-of-sample explanatory power by 0.486.

Figure 2 plots observed against predicted $\hat{\sigma}$. The decomposition captures the ranking and cross-subject variation closely; the largest errors occur in the upper tail.

The likelihood guarantees that $\hat{\sigma}$ responds to within-sample choice deviations; it does not guarantee that coefficients learned elsewhere predict its level and ranking in Study 2. The exercise therefore provides the paper's strongest evidence that model-relative fit contains stable information beyond inconsistency alone. The claim remains bounded: the prediction transports across populations performing the same class of portfolio task,

Table 6: Out-of-Sample Prediction of $\hat{\sigma}$ in Study 2

	Forecast accuracy			Mincer–Zarnowitz calibration		
	R^2_{OOS}	Pearson r	Spearman ρ	Slope	Intercept	p-value
Panel A: Money-metric decomposition model						
Option A: pure held-out prediction	0.816	0.913	0.930	0.961	−0.014	0.001
Option B: re-centred prediction	0.832	0.913	0.930	0.961	0.006	0.669
Panel B: Varian-only benchmark, same refined sample						
Option A: pure held-out prediction	0.330	0.585	0.772	0.930	0.026	0.010
Option B: re-centred prediction	0.340	0.585	0.772	0.930	0.011	0.473

Notes. The table evaluates prediction of $\hat{\sigma}$ in held-out Study 2. The money-metric decomposition model is estimated on Studies 1, 3, and 4 and regresses $\hat{\sigma}$ on the Varian inefficiency index, the DA-CRRA misspecification residual, and study fixed effects for Studies 3 and 4. The Varian-only benchmark is estimated on the same refined sample and excludes the DA-CRRA misspecification residual. Both models use 235 training observations and 535 Study 2 test observations. Option A assigns Study 2 the omitted Study 1 intercept. Option B adds the mean Study 2 residual from Option A to all predicted values. The level shift is −0.020 for the decomposition model and 0.017 for the Varian-only benchmark. $R^2_{\text{OOS}} = 1 - \text{SSE}/\text{SST}$, where SST is computed relative to the Study 2 mean of observed $\hat{\sigma}$. The Mincer–Zarnowitz p-value is from the joint test of zero intercept and unit slope in the regression of observed $\hat{\sigma}$ on predicted $\hat{\sigma}$, estimated in the Study 2 test sample with robust standard errors. The slope and intercept columns report the estimated Mincer–Zarnowitz calibration coefficients; significance stars are omitted because the relevant calibration test is the joint null of zero intercept and unit slope, reported in the final column.

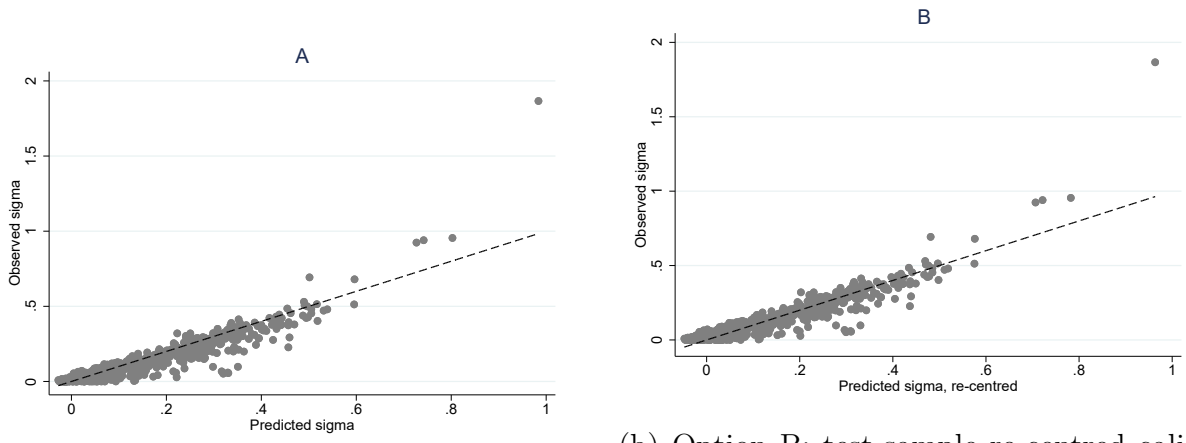
not yet across elicitation designs or to welfare-relevant outcomes outside the experiment.

7 Validation and Robustness

7.1 Selection and Estimation Robustness

A first concern is selection: if the DA–RUM converged only for highly consistent subjects, the main results would describe an unusually precise subset. Table 7 relates numerical convergence to GARP in all 1,650 subjects.

Convergence is positively related to GARP, but modestly over the observed range. The linear probability model and probit marginal effect are both approximately 0.21 for a unit change in GARP; a 0.10 decline in GARP therefore predicts only about a 2 percentage point decline in convergence. Convergence is not confined to near-perfectly



(a) Option A: pure held-out prediction

(b) Option B: test-sample re-centred calibration

Figure 2: Observed and Predicted $\hat{\sigma}$ in Study 2

Notes. The figure plots observed $\hat{\sigma}$ against predicted $\hat{\sigma}$ for held-out Study 2 subjects under the money-metric decomposition model. The model is estimated on Studies 1, 3, and 4 only. The left panel reports strict held-out predictions, assigning Study 2 the omitted Study 1 intercept. The right panel adds the mean Study 2 residual from the strict prediction, shifting predictions by -0.020. The dashed line is the 45-degree line.

consistent subjects.

The relationship is insignificant in the denser 50-decision Study 1 sample, although that sample is small. Appendix Table A6 likewise finds small, insignificant GARP gaps between converged and non-converged subjects in Studies 1–3; only Study 4 shows a significant gap. Selection is present, but not uniformly strong.

The main mappings also survive broader samples that relax convergence, the $\hat{a} < 15$ trim, and approximation-quality restrictions (Appendix D); exclusion of the least precise first-stage $\hat{\sigma}$ estimates (Appendix E); and inverse-hyperbolic-sine and quantile specifications (Appendix Table A8). The evidence is therefore not driven by a narrow estimation sample or by the upper tail of $\hat{\sigma}$.

7.2 Separating Inconsistency from Model Misspecification

The observational data cannot independently manipulate consistency and model spanning. The simulation does so while preserving Study 1's budget sequences and applying the same DA–RUM estimator used in the empirical analysis.

The correctly specified benchmark generates choices from DA-CRRA with $(a, \rho) = (1.47, 0.21)$, the Study 1 medians. Stochastic inconsistency is introduced through a normal perturbation to the log consumption ratio, $\ell = \log(x_1/x_2)$, with scale τ. We calibrate

Table 7: GARP Consistency and DA–RUM Convergence

	(1) Pooled Probit	(2) Pooled OLS	(3) Probit (Study 1)
GARP	0.535**	0.211**	-2.876
	(0.231)	(0.091)	(2.189)
Study 2	0.136	0.054	–
	(0.188)	(0.075)	
Study 3	-0.007	-0.003	–
	(0.204)	(0.082)	
Study 4	0.301	0.118	–
	(0.204)	(0.081)	
Constant	-0.473*	0.313***	2.719
	(0.284)	(0.113)	(2.071)
Observations	1,650	1,650	47
Pseudo R^2	0.0043	–	0.0305
R^2	–	0.0058	–
Wald χ^2 / F	9.72	2.48	1.73
Average Marginal Effect of GARP			
GARP	0.210**		-1.110
	(0.090)		(0.794)

Notes. Dependent variable: indicator for convergence of the individual-level DA–RUM estimation. Column (1) reports pooled probit estimates with study fixed effects; the average marginal effect of GARP is reported below the main estimates. Column (2) reports the corresponding linear probability model. Column (3) reports a probit specification for Study 1 only, where each subject completes 50 decisions. Robust standard errors in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

τ separately within each data-generating process to CCEI targets of 0.97 and 0.88.[2] The perturbed ratio is mapped back to a feasible bundle, so every simulated choice satisfies its budget.

Two comparison data-generating processes vary spanning. Expected utility sets $a = 1$ and is nested within DA. A token-guarantee heuristic first secures $c_t = \min\{10, 1/(p_{1t} + p_{2t})\}$ in both states and allocates residual wealth to the cheaper asset. This disciplined but non-smooth rule is outside DA-CRRA. We apply the same log-ratio perturbation and CCEI targets to every process.

Table 8 shows that, under correct DA specification, lowering CCEI from 0.97 to 0.88 raises mean $\hat{\sigma}$ from 0.034 to 0.121. The curvature estimate changes only modestly,

[2]Observed choices are generated according to

$$\ell_t = \ell_t^* + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \tau^2),$$

where τ controls stochastic inconsistency.

Table 8: Simulation Results: Parameter Recovery and Convergence

Scenario	$\hat{\rho}$	$\hat{a}$	$\hat{\sigma}$	Convergence	CCEI
A1: DA (Low Noise)	0.288	1.227	0.034	39/42	0.97
A2: DA (High Noise)	0.277	1.076	0.121	36/41	0.88
D1: Heuristic (Low Noise)	0.545	1.000	0.217	39/44	0.97
D2: Heuristic (High Noise)	0.419	1.010	0.279	26/41	0.88
C1: EU (Low Noise)	0.199	1.020	0.042	39/43	0.97
C2: EU (High Noise)	0.173	1.057	0.106	36/40	0.88

Notes. The calibration uses Study 1 parameter values. DA scenarios set $a = 1.47$ and $\rho = 0.21$; EU scenarios set $a = 1$ and $\rho = 0.21$. Within each data-generating process, τ is calibrated to CCEI targets of 0.97 and 0.88. Estimates use converged solutions; convergence rates are reported.

although $\hat{a}$ also responds. Dispersion absorbs much, but not necessarily all, of the added variation.

The heuristic comparison isolates model spanning. At CCEI 0.97, mean $\hat{\sigma}$ is 0.217 under the heuristic but 0.034 under DA. Since consistency is matched, the difference reflects the inability of DA-CRRA to reproduce the heuristic. At CCEI 0.88, heuristic dispersion rises further to 0.279. The recovered preference parameters also move, confirming that a stochastic residual does not fully insulate preferences from misspecification.

The nested EU cases rule out a trivial "different DGP" explanation. Their $\hat{\sigma}$ estimates, 0.042 and 0.106, closely track the corresponding DA cases. Dispersion rises sharply when the DA-CRRA family cannot span the rule, not whenever the generating parameters differ.

Within this calibrated environment, $\hat{\sigma}$ therefore responds to both inconsistency and lack of model spanning. The exercise illustrates the mechanism; it does not establish the quantitative share of empirical dispersion due to either source.

7.3 External Validation: Cognitive Ability

We next ask whether an external measure of decision-making capacity is associated with preferences, dispersion, or both.

Study 2 measures Cognitive Reflection Test performance; Study 4 measures IQ. We standardise each measure within study, so cog_z represents a one-standard-deviation increase in the relevant instrument. Studies 1 and 3 lack comparable measures.

Table 9 finds no significant association with $\hat{\rho}$ or $\hat{a}$. A one-standard-deviation increase in cognitive ability is associated with a 0.031 reduction in $\hat{\sigma}$. Adding GARP reduces this coefficient to -0.007, which is insignificant at 5 percent but marginally significant at 10 percent; the Study 4 interaction is negligible.

The attenuation is informative but not a causal mediation result. It shows that most of the cognitive gradient in dispersion overlaps with variation captured axiomatically by GARP. This extends the pattern in Andersson et al. (2020) to continuous budgets and is consistent with limits to information processing affecting approximate preference implementation (Enke, 2024). In the paper's decomposition, cognition is associated primarily with the inconsistency-related component of dispersion, not with recovered tastes.

Table 9: DA–RUM Parameters, GARP, and Cognitive Ability

	(1) $\hat{\rho}_i$	(2) $\hat{a}_i$	(3) $\hat{\sigma}_i$	(4) $\hat{\sigma}_i$	(5) $\hat{\sigma}_i$
Cognitive Ability (z)	-0.019	0.097	-0.031***	-0.007*	-0.007*
	(0.013)	(0.072)	(0.006)	(0.003)	(0.004)
Study 4	0.026	-0.881***	0.013	0.011	0.011
	(0.030)	(0.089)	(0.016)	(0.012)	(0.012)
GARP				-1.037***	-1.037***
				(0.043)	(0.043)
Study 4 × *cog_z*					0.001
					(0.010)
Observations	777	777	777	777	777

Notes. Cognitive ability is standardised separately within Studies 2 and 4. Studies 1 and 3 do not contain comparable cognitive measures and are excluded. All specifications include a Study 4 indicator. Column (5) includes the interaction between Study 4 and cognitive ability. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

8 Why the Distinction Matters for Preference Rankings

Applied work often ranks people or groups by recovered preferences while treating dispersion as a nuisance. Our results imply that these are joint outputs of the fitted model: the same value of a preference index can be supported by choice patterns with very different residual dispersion. To show that this qualification is empirically relevant, we examine $\hat{\sigma}$ across the distribution of DA–RUM implied local risk aversion, $r(1)$.

We summarise recovered risk attitudes using the DA–RUM implied local risk-aversion index $r(1)$:

$$r(1) \approx \frac{\hat{a}-1}{\hat{a}+1} + \hat{\rho}\frac{2\hat{a}}{(\hat{a}+1)^2}.$$

Following Choi et al. (2007), the index combines curvature and disappointment aversion into a scalar measure of local risk attitudes around certainty.

Figure 3 shows a negative gradient. Subjects with lower recovered $r(1)$ have higher average dispersion; subjects in the upper part of the $r(1)$ distribution have comparatively low $\hat{\sigma}$. High recovered risk aversion is therefore not concentrated among the highest-dispersion fits.

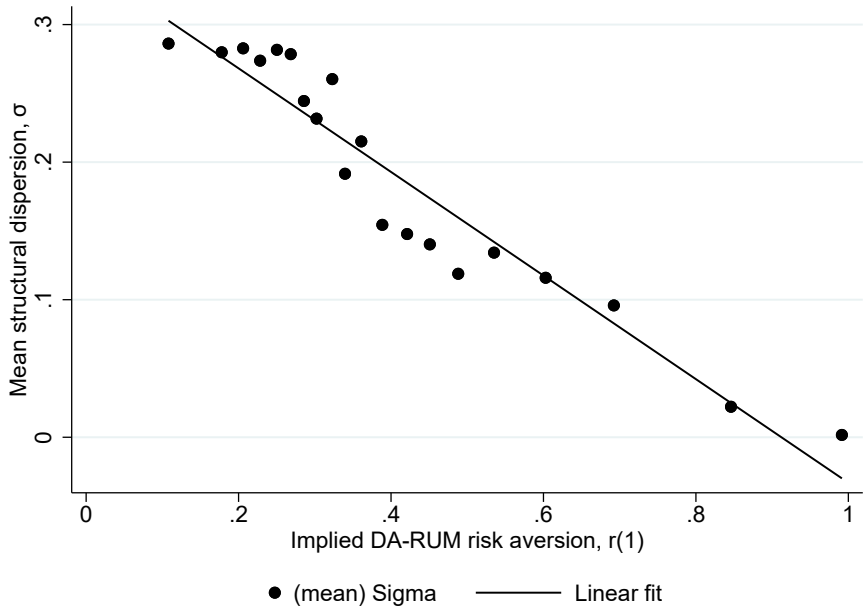


Figure 3: Structural Dispersion across Recovered Risk Attitudes

Notes. The figure plots mean estimated structural dispersion, $\hat{\sigma}$, across bins of the DA-RUM implied local risk-aversion index $r(1)$. The fitted line summarises the unconditional association. Appendix Table A12 reports regressions including study fixed effects.

Appendix Table A11 quantifies the pattern. Mean $\hat{\sigma}$ falls from 0.281 in the lowest $r(1)$ quartile to 0.074 in the highest; the share in the top dispersion quartile falls from 43.6 to 5.8 percent. With study fixed effects, average $\hat{\sigma}$ is lower by 0.042, 0.123, and 0.216 in successive risk-aversion quartiles relative to the lowest. A continuous specification yields a coefficient of -0.443, with no significant quadratic term (Appendix Table A12).

The gradient requires caution. Appendix A shows that menu normalisation substantially reduces the raw curvature–dispersion scale channel, but $r(1)$ and $\hat{\sigma}$ are still jointly recovered and need not be mechanically independent. Nor is $\hat{\sigma}$ a standard error that can be used directly to weight preference estimates. The result is descriptive: residual dispersion is unevenly distributed across recovered preference types.

That unevenness is economically relevant. Comparisons of average preferences across groups can coincide with differences in consistency or model fit, and a preference ranking

alone conceals that distinction. Reporting $\hat{\sigma}$ alongside recovered preferences exposes the issue; the MMI decomposition then diagnoses whether the accompanying dispersion is associated with inconsistency, DA-CRRA lack of fit, or both.

9 Conclusion

Structural estimation divides observed choice into preferences and a stochastic residual. Our evidence shows that this division is model-dependent. In four portfolio-choice studies, the DA–RUM dispersion parameter is associated with both revealed-preference inconsistency and the additional money-metric loss from imposing DA-CRRA. The misspecification component remains predictive among axiomatically consistent subjects.

The held-out exercise gives the result empirical bite. A mapping estimated in three studies predicts dispersion in a fourth with an out-of-sample R^2 of 0.816; removing DA-CRRA misspecification lowers this to 0.330. The simulation reproduces the mechanism: inconsistency raises dispersion under correct specification, and a non-spanned but structured rule raises it further at matched consistency. Cognition aligns mainly with the inconsistency-related component.

The implication is not that $\hat{\sigma}$ should replace a standard error or serve as a model-free index of rationality. It is that a stochastic residual carries information about both implementation and the adequacy of the fitted preference family. Preference estimates and dispersion should therefore be reported jointly, and cross-group differences in either should be interpreted only after asking whether the same deterministic model fits the groups equally well.

Our evidence is specific to DA-CRRA estimation in a common portfolio environment. Establishing the same decomposition across preference families and elicitation designs is the next step. Within this setting, however, the conclusion is clear: what structural estimation calls noise is partly a property of behaviour and partly a property of the model used to describe it.

References

Afriat, Sydney N (1967), "The construction of utility functions from expenditure data." *International Economic Review*, 8, 67–77.

Andersen, Steffen, Glenn W Harrison, Morten I Lau, and E Elisabet Rutström (2008), "Eliciting risk and time preferences." *Econometrica*, 76, 583–618.

Andersson, Ola, Håkan J Holm, Jean-Robert Tyran, and Erik Wengström (2020), "Robust inference in risk elicitation tasks." *Journal of Risk and Uncertainty*, 61, 195–209.

Apesteguia, Jose and Miguel A Ballester (2015), "A measure of rationality and welfare." *Journal of Political Economy*, 123, 1278–1310.

Apesteguia, Jose and Miguel A Ballester (2018), "Monotone stochastic choice models: The case of risk and time preferences." *Journal of Political Economy*, 126, 74–106.

Chapman, Jonathan and Geoffrey Fisher (2025), "Preference elicitation: common methods and potential pitfalls." *Handbook of Experimental Methodology*, 1, 25–80.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv (2007), "Consistency and heterogeneity of individual behavior under uncertainty." *American Economic Review*, 97, 1921–1938.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman (2014), "Who is (more) rational?" *American Economic Review*, 104, 1518–50.

Echenique, Federico, Taisuke Imai, and Kota Saito (2023), "Approximate expected utility rationalization." *Journal of the European Economic Association*, 21, 1821–1864.

Enke, Benjamin (2024), "The cognitive turn in behavioral economics." *Unpublished manuscript. https://benjamin-enke.com.*

Gul, Faruk (1991), "A theory of disappointment aversion." *Econometrica*, 59, 667–686.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill (2018), "Parametric recoverability of preferences." *Journal of Political Economy*, 126, 1558–1593.

Lau, Morten I and Hong Il Yoo (2025), "Structural estimation of higher order risk preferences." *Econometrica*, 93, 1855–1883.

McFadden, Daniel (1972), "Conditional logit analysis of qualitative choice behavior." *In P. Zarembka (ed.), Frontiers in Econometrics*, 105–142.

Moghaddasi Kelishomi, Ali and Daniel Sgroi (2024), "A field study of donor behaviour in the iranian kidney market." *European Economic Review*, 170, 104887.

Polisson, Matthew, John K-H Quah, and Ludovic Renou (2020), "Revealed preferences over risk and uncertainty." *American Economic Review*, 110, 1782–1820.

Varian, Hal R (1982), "The nonparametric approach to demand analysis." *Econometrica*, 50, 945–973.

Varian, Hal R (1990), "Goodness-of-fit in optimizing models." *Journal of Econometrics*, 46, 125–140.

Wilcox, Nathaniel T (2011), "Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk." *Journal of Econometrics*, 162, 89–104.

Appendix

A Menu-Specific Normalisation and the Curvature-Noise Relationship

A potential concern in Fechner-type random utility models is that the estimated dispersion parameter may partly reflect the cardinal scale of the utility index. In an unnormalised specification, choice probabilities depend on utility differences divided by σ. If changes in the curvature parameter mechanically rescale utility differences, then the value of σ required to rationalise a given stochastic choice pattern will also change mechanically with curvature.

This appendix quantifies the role played by the menu-specific normalisation in the DA–RUM. Using representative Study 1 budgets, we compare the variation in the implied dispersion parameter generated by changes in curvature before and after normalisation. The exercise therefore measures how much of the raw curvature-dispersion scale channel is removed by expressing utility differences relative to the within-menu utility spread.

We use five representative Choi budgets from the Study 1 grid data, chosen to span different price ratios. For each budget, we compare two benchmark allocations. The first is the safe allocation on the 45-degree line,

$$S_t = (s_t, s_t), \qquad s_t = \frac{x_{10,t}x_{20,t}}{x_{10,t} + x_{20,t}},$$

where $x_{10,t}$ and $x_{20,t}$ are the horizontal and vertical intercepts of the budget line. The second is the risky corner allocation on the cheaper-security axis:

$$R_t = \begin{cases} (x_{10,t}, 0), & \text{if } x_{10,t} \geq x_{20,t}, \\ (0, x_{20,t}), & \text{if } x_{20,t} > x_{10,t}. \end{cases}$$

The comparison between S_t and R_t provides a transparent benchmark for measuring how curvature changes the utility separation between a hedged allocation and a risky allocation. Applying the same comparison across the curvature grid shows how strongly the implied dispersion scale changes before and after menu-specific normalisation.

For each value of ρ on a grid from 0 to 1, holding $a = 1.47$, we compute the raw DA utility difference

$$\Delta U_t(\rho, a) = U(S_t; \rho, a) - U(R_t; \rho, a),$$

using the same DA-CRRA utility specification as in the main estimation. We then compute the menu-specific scale

$$S_{it}(\rho, a) = sd\{U(h; \rho, a) - \bar{U}_t(\rho, a) : h \in G_{it}\},$$

where G_t is the discretised menu associated with budget t. The corresponding menu-normalised utility difference is

$$\widetilde{\Delta U}_t(\rho, a) = \frac{\Delta U_t(\rho, a)}{S_t(\rho, a)}.$$

To translate the utility-gap comparison into the noise parameter, we compute the value of σ that would rationalise a fixed choice probability between the two benchmark allocations. Let $p = 0.90$ denote the probability of choosing the utility-favoured benchmark allocation. In a binary logit comparison, the associated log-odds are

$$\ell = \log\left(\frac{p}{1-p}\right).$$

Thus, the implied value of σ in the unnormalised specification is

$$\sigma_t^{raw}(\rho, a) = \frac{|\Delta U_t(\rho, a)|}{\ell},$$

while the corresponding value under menu-specific normalisation is

$$\sigma_t^{norm}(\rho, a) = \frac{|\widetilde{\Delta U}_t(\rho, a)|}{\ell} = \frac{|\Delta U_t(\rho, a)|}{S_t(\rho, a)\ell}.$$

The absolute value is used because the exercise concerns the amount of noise required to rationalise a fixed probability of choosing the utility-favoured bundle, rather than the direction of preference between the two benchmark allocations.

Figure A1 plots $\sigma_t^{raw}(\rho, a)$ and $\sigma_t^{norm}(\rho, a)$ across the ρ grid. Without normalisation, the implied noise parameter changes sharply with curvature. After normalising by $S_t(\rho, a)$, the implied value of σ is substantially flatter.

Table A1 quantifies this effect. For each representative budget, we compute the range of implied σ over the ρ grid before and after menu-specific normalisation:

$$Range_t^{raw} = \max_{\rho} \sigma_t^{raw}(\rho, a) - \min_{\rho} \sigma_t^{raw}(\rho, a),$$

and

$$Range_t^{norm} = \max_{\rho} \sigma_t^{norm}(\rho, a) - \min_{\rho} \sigma_t^{norm}(\rho, a).$$

The reduction measure is

$$100 \times \left(1 - \frac{Range_t^{norm}}{Range_t^{raw}}\right).$$

Across the five representative budgets, menu-specific normalisation reduces the range of implied σ by approximately 75% on average.

This diagnostic shows that the normalisation substantially mitigates the mechanical scale channel through which changes in curvature would otherwise be loaded into σ. It does not imply that $\hat{\rho}$, $\hat{a}$, and $\hat{\sigma}$ are fully orthogonal, since $S_t(\rho, a)$ itself depends on the preference parameters. Rather, it illustrates that $\hat{\sigma}$ is estimated relative to the local utility dispersion of the relevant menu, instead of raw CRRA utility units.

B Exogenous Uniform-Grid Robustness

The baseline DA–RUM estimation uses the Polisson et al. (2020) GRID construction, in which the finite menu is generated from the set of consumption levels observed in a

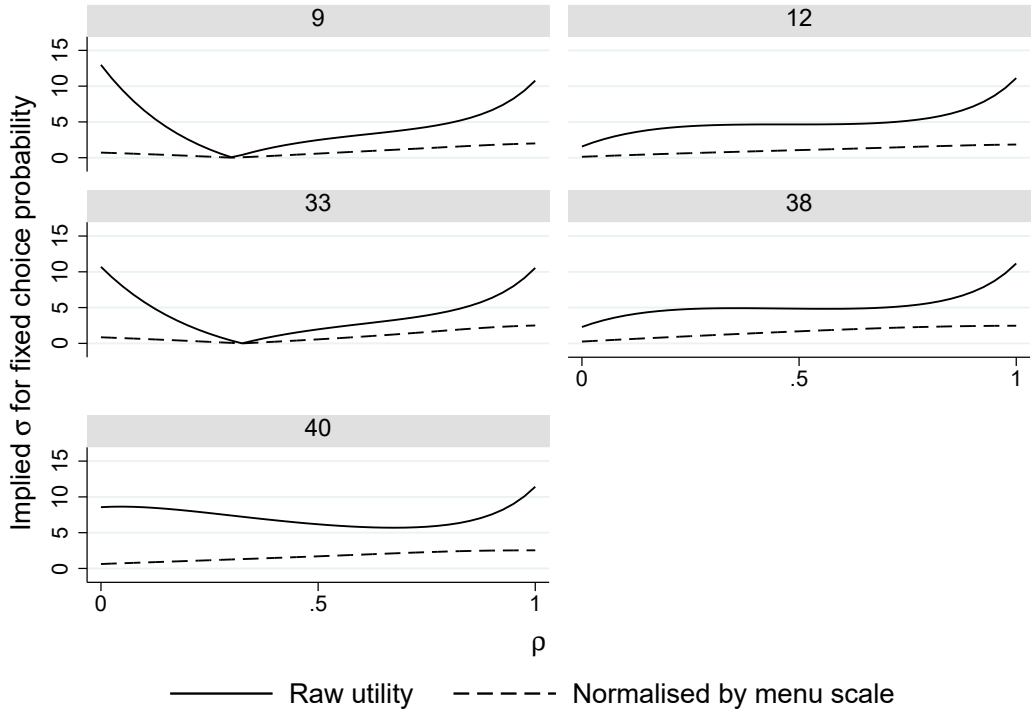


Figure A1: Raw and Menu-Normalised Implied σ Across Representative Budgets

Notes. The figure plots the value of σ required to rationalise a fixed 0.90 choice probability as ρ varies across five representative Choi budgets. The solid line uses raw DA utility differences, while the dashed line divides utility differences by the within-menu utility dispersion $S_{it}(\rho, a)$. Menu-specific normalisation substantially reduces the curvature-induced variation in implied σ.

subject's choices. This construction is natural for revealed-preference testing, but it implies that the menu entering the likelihood is partly choice-dependent. To assess whether the main RP–RUM mapping is driven by this feature, we re-estimate the DA–RUM in Studies 3 and 4 using an exogenous uniform discretisation of each budget line.

Specifically, each budget line is discretised into 100 equally spaced alternatives.[3] The observed chosen allocation is appended to the menu to ensure that the actual choice is available in the likelihood. This creates an alternative likelihood menu that no longer depends on the subject's realised consumption coordinates across other rounds.

The resulting estimates differ from the baseline GRID estimates, as expected when the likelihood menu changes. On the common converged and trimmed sample, the uniform-grid estimates correlate with the baseline estimates at 0.740 for $\hat{\rho}$, 0.662 for $\hat{a}$, and 0.559 for $\hat{\sigma}$. The alternative menu construction therefore changes the cardinal scale and some of the cross-subject ranking of the estimates, particularly for $\hat{\sigma}$, while preserving substantial common signal across the two menu definitions.

[3] As an additional check, we repeated the exogenous-grid estimation in Study 3 using 200 alternatives per budget line. The resulting estimates are virtually identical to the 100-point version, with correlations of 1.000 for $\hat{\rho}$, 0.999 for $\hat{a}$, and 0.998 for $\hat{\sigma}$.

Table A1: Reduction in Curvature-Induced Variation in Implied σ

Budget Round	x_{10}	x_{20}	Range raw	Range normalised	Reduction (%)
9	33	88	13	2	85
12	70	52	10	2	82
33	67	24	11	3	77
38	56	73	9	2	75
40	85	83	6	2	67
Mean					75.2

Notes. The table reports the range of implied σ over the ρ grid before and after menu-specific normalisation. The implied σ is calculated as the absolute benchmark-bundle utility gap divided by the fixed log-odds corresponding to a 0.90 choice probability. Reduction is computed as $100 \times [1 - \text{Range}_{norm}/\text{Range}_{raw}]$.

More importantly, the main empirical relationships are qualitatively unchanged. As shown in Table A2, structural noise remains negatively related to revealed-preference consistency and positively related to both Varian inefficiency and the DA-CRRA misspecification residual. The magnitudes of the coefficients are not directly comparable to the baseline estimates because the scale of $\hat{\sigma}$ changes with the likelihood menu. The robustness exercise therefore focuses on the sign and significance of the mapping rather than the cardinal size of the coefficients. These results indicate that the relationship between $\hat{\sigma}$ and the revealed-preference diagnostics is not an artefact of the choice-dependent GRID construction.

C Split-Sample Validation of Dispersion and Consistency

A potential concern with the main regressions in Table 4 is that both the DA–RUM dispersion parameter and the revealed-preference indices are computed from the same set of choices. The relationship between $\hat{\sigma}$ and GARP could therefore partly reflect mechanical overlap in the underlying observations. To assess this concern, we use the 50-round Choi et al. (2007) data and split each subject's choices into odd and even rounds. We then estimate the DA–RUM separately on one half of the data and compute GARP consistency on the complementary half.

As a first check on the reliability of the split-sample estimates, we compare DA–RUM parameters estimated separately on odd and even rounds in Study 1. Among subjects for whom both split estimations converge, the odd–even correlations are high: 0.815 for $\hat{\rho}$, 0.916 for $\hat{a}$, and 0.749 for $\hat{\sigma}$. Although the common converged sample is small, these correlations indicate that the estimated dispersion parameter has substantial within-subject persistence across non-overlapping sets of choices, rather than being a purely round-specific fitting residual.

We then use the same split design to address mechanical overlap between structural

Table A2: Exogenous Uniform-Grid Robustness: Studies 3 and 4

	RP consistency mapping $\hat{\sigma}$	Misspecification mapping $\hat{\sigma}$
GARP consistency	-6.961^{***}	
	(1.819)	
Misspecification residual		8.597^{***}
		(1.630)
Varian inefficiency		17.224^{***}
		(5.229)
Study fixed effect	Yes	Yes
Observations	207	202

Notes. The table reports robustness estimates using an exogenous 100-point uniform discretisation of each budget line for Studies 3 and 4. The observed chosen allocation is appended to each menu. Column (1) replicates the specification in column (3) of Table 4, replacing the baseline Polisson GRID with the exogenous uniform grid. Column (2) replicates the specification in column (3) of Table 5. Coefficients are not directly comparable in magnitude to the baseline estimates because the scale of $\hat{\sigma}$ changes with the likelihood menu. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

noise and revealed-preference consistency. Specifically, we estimate $\hat{\sigma}$ on one half of the choices and compute GARP consistency on the complementary half.

Table A4 reports the results. The coefficient on complementary-split GARP is negative in both directions and statistically significant in the pooled specification. Thus, subjects whose choices are more consistent in one half of the experiment also exhibit lower structural noise when $\hat{\sigma}$ is estimated using the other half. Although this exercise is based on a relatively small sample and each split contains only 25 choices, it shows that the mapping between structural noise and revealed-preference consistency is not driven solely by using the same choice observations to construct both objects.

Table A3: Split-Half Reliability of DA–RUM Estimates

Parameter	Odd–even correlation
$\hat{\rho}$	0.815
$\hat{a}$	0.916
$\hat{\sigma}$	0.749
Observations	18

Notes. The table reports correlations between DA–RUM estimates obtained separately from odd and even rounds in Study 1. The sample is restricted to subjects for whom both split-sample estimations converge. Each split contains 25 decisions of the 50-round Study 1 data.

Table A4: Split-Sample Relationship Between Structural Dispersion and Revealed-Preference Consistency

	$\hat{\sigma}^{odd}$ on $GARP^{even}$	$\hat{\sigma}^{even}$ on $GARP^{odd}$	Pooled split sample
Complementary-split GARP	-0.593^{***}	-0.703^{*}	-0.626^{***}
	(0.105)	(0.374)	(0.137)
Direction fixed effect			0.019
			(0.026)
Observations	29	24	53
R^2	0.393	0.158	0.256

Notes. The table reports split-sample regressions using Choi et al. (2007). The DA–RUM is estimated separately on odd and even rounds. Revealed-preference consistency is computed on the complementary half of the data. The pooled specification stacks both split directions and includes a direction fixed effect. Standard errors are robust in columns 1-2 and clustered by subject in column 3.

D Robustness to Convergence, Trimming, and Approximation Restrictions

Table A5 examines whether the main results are driven by convergence filters, the trimming of extreme disappointment-aversion estimates, or the exclusion of weakly approximated Varian indices. Panel A shows that the negative relationship between structural noise and GARP consistency remains large and highly significant when the convergence restriction is removed and when the $\hat{a} < 15$ trim is also relaxed. The coefficient on GARP changes from -1.041 in the baseline specification to approximately -0.91 in the broader samples.

Panel B performs the same exercise for the decomposition of structural noise into Varian inefficiency and misspecification. Both components remain positive, large, and highly significant across all specifications. The misspecification coefficient is particularly stable, ranging from 2.278 to 2.349 across the alternative samples. The coefficient on Varian inefficiency also remains strongly positive. These results indicate that the central mapping between $\hat{\sigma}$, revealed-preference inconsistency, and model-relative misspecification is not driven by the convergence filter, the trimming threshold, or the approximation-quality restriction.

E Robustness to First-Stage Precision of $\hat{\sigma}$

The main second-stage regressions use estimated individual-level dispersion parameters. A concern is that some values of $\hat{\sigma}$ may be weakly identified in the first-stage DA–RUM estimation. To assess whether such observations drive the results, Table A7 repeats the main specifications after excluding observations with the largest first-stage standard errors of $\hat{\sigma}$.

Table A5: Robustness to Convergence, Trimming, and Approximation Restrictions

	Baseline	Drop convergence	Drop trim	Full sample
Panel A: Structural dispersion and GARP consistency				
GARP	-1.041^{***}	-0.913^{***}	-0.921^{***}	
	(0.040)	(0.041)	(0.041)	
Observations	906	1,313	1,360	
Panel B: Structural dispersion, Varian inefficiency, and misspecification				
Varian inefficiency	1.877^{***}	1.867^{***}	1.843^{***}	2.226^{***}
	(0.231)	(0.173)	(0.172)	(0.065)
Misspecification	2.282^{***}	2.349^{***}	2.338^{***}	2.278^{***}
	(0.231)	(0.112)	(0.111)	(0.093)
Observations	770	1,083	1,130	1,360

Notes. Panel A reports the coefficient on GARP consistency from regressions of $\hat{\sigma}$ on GARP and study fixed effects. The baseline column corresponds to the main specification in Table 4, column (3), which restricts the sample to converged DA–RUM estimates with $\hat{a} < 15$. The second column removes the convergence restriction but retains the $\hat{a} < 15$ trim. The third column removes both the convergence restriction and the $\hat{a} < 15$ trim. Panel B reports the coefficients on Varian inefficiency and the misspecification residual from the analogue of Table 5, column (3). The baseline column restricts the sample to converged DA–RUM estimates with $\hat{a} < 15$ and excludes weakly approximated Varian indices. The second column drops the convergence restriction, the third additionally drops the $\hat{a} < 15$ trim, and the fourth uses the full sample including weakly approximated Varian indices. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A shows that the negative relationship between structural noise and GARP consistency remains large and statistically significant after removing the top 5% and top 10% most imprecisely estimated $\hat{\sigma}$ values. Panel B shows the corresponding robustness checks for the decomposition into Varian inefficiency and misspecification. Both components remain positive and statistically significant across the restricted samples. These results suggest that the main findings are not driven by observations with weakly identified first-stage dispersion estimates.

F Alternative Specifications for the Distribution of $\hat{\sigma}$

Table A8 reports robustness checks for the main structural-dispersion regressions. The $asinh(\hat{\sigma})$ specifications replace $\hat{\sigma}$ with the inverse hyperbolic sine transformation, which is log-like for positive values but defined at zero. Q_{25}, Q_{50}, and Q_{75} report quantile regressions at the 25th, 50th, and 75th percentiles of $\hat{\sigma}$.

Table A6: Mean GARP Consistency by DA–RUM Convergence Status

Study	N_C	Mean GARP C	N_{NC}	Mean GARP NC	Difference
Study 1	24	0.916	23	0.954	−0.038
Study 2	654	0.885	528	0.875	0.010
Study 3	105	0.975	98	0.984	−0.010
Study 4	133	0.881	85	0.797	0.084***

Notes. C denotes subjects whose individual-level DA–RUM estimates converged, and NC denotes subjects whose estimates did not converge. Difference is mean GARP for converged subjects minus mean GARP for non-converged subjects. p-values are from robust regressions of GARP on a convergence indicator, estimated separately by study. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

G Alternative Measures of Consistency for Table 5

The consistent-sample result in Table 5 shows that the DA-CRRA misspecification residual predicts $\hat{\sigma}$ even among subjects for whom revealed-preference inconsistency is absent under the baseline criterion. Table A9 examines whether this result depends on the definition of consistency. We repeat the exercise using exact GARP consistency, as well as very high fGARP and EU thresholds. In all cases, the misspecification residual remains positive and statistically significant. This indicates that the relationship between DA-CRRA lack of fit and structural dispersion is not driven by a particular revealed-preference consistency criterion.

Table A7: Robustness to First-Stage Precision of $\hat{\sigma}$

	Baseline	Exclude top 5%	Exclude top 10%
Panel A: Structural noise and GARP consistency			
GARP	-1.041^{***}	-0.966^{***}	-0.910^{***}
	(0.040)	(0.036)	(0.036)
Observations	906	861	816
Panel B: Structural noise, Varian inefficiency, and misspecification			
Varian inefficiency	1.877^{***}	2.226^{***}	2.209^{***}
	(0.231)	(0.099)	(0.109)
Misspecification	2.282^{***}	1.746^{***}	1.667^{***}
	(0.231)	(0.056)	(0.064)
Observations	770	732	693

Notes. Panel A reports the coefficient on GARP consistency from regressions of $\hat{\sigma}$ on GARP and study fixed effects. Panel B reports the coefficients on Varian inefficiency and the DA-CRRA misspecification residual from regressions of $\hat{\sigma}$ on both components and study fixed effects. The baseline columns correspond to the main specifications in Tables 4 and 5. The second and third columns exclude observations with the largest first-stage standard errors of $\hat{\sigma}$. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Distributional Robustness of the Structural Dispersion Results

	Panel A: GARP consistency			*Panel B: Varian decomposition*		
	$asinh(\hat{\sigma})$	Q_{25}	Q_{50}	$asinh(\hat{\sigma})$	Q_{25}	Q_{50}
GARP	-0.994^{***}	-1.028^{***}	-1.183^{***}			
	(0.035)	(0.030)	(0.035)			
Varian inefficiency				2.285^{***}	2.203^{***}	2.230^{***}
				(0.033)	(0.038)	(0.030)
DA-CRRA misspecification				2.013^{***}	1.633^{***}	1.810^{***}
				(0.101)	(0.063)	(0.044)
Observations	906	906	906	899	899	899
R^2 / pseudo-R^2	0.570	0.392	0.443	0.910	0.658	0.734

	Panel A continued	*Panel B continued*	
	Q_{75}	Q_{75}	Consistent subjects: $asinh(\hat{\sigma})$
GARP	-1.206^{***}		
	(0.044)		
Varian inefficiency		2.399^{***}	
		(0.027)	
DA-CRRA misspecification		2.041^{***}	2.197^{***}
		(0.037)	(0.419)
Observations	906	899	208
R^2 / pseudo-R^2	0.444	0.765	0.806

Notes. The table reports robustness checks for the main structural-dispersion regressions. The $asinh(\hat{\sigma})$ specifications replace $\hat{\sigma}$ with the inverse hyperbolic sine transformation, which is log-like for positive values but defined at zero. Q_{25}, Q_{50}, and Q_{75} report quantile regressions at the 25th, 50th, and 75th percentiles of $\hat{\sigma}$. Panel A corresponds to the GARP specification in Table 4. Panel B corresponds to the Varian inefficiency and DA-CRRA misspecification specification in Table 5. The final column restricts the sample to subjects with zero Varian inefficiency, Varian inefficiency $\leq 10^{-12}$, corresponding to the consistent-subject specification. All specifications include study fixed effects. Standard errors are in parentheses. For the $asinh(\hat{\sigma})$ specifications, robust standard errors are reported. For the quantile regressions, bootstrap standard errors with 500 replications are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Misspecification and Structural Dispersion Among Consistent and Highly Consistent Subjects

	GARP-consistent $\hat{\sigma}$	Baseline consistent $\hat{\sigma}$	High fGARP $\hat{\sigma}$	High EU $\hat{\sigma}$
Misspecification residual	2.695***	2.678***	1.647***	1.570***
	(0.713)	(0.697)	(0.220)	(0.245)
Consistency definition	GARP = 1	Baseline consistent	fGARP > 0.998	EU > 0.9968
Study fixed effects	Yes	Yes	Yes	Yes
Observations	189	208	47	45

Notes. The table reports robustness checks for the consistent-subject result in Table 5. The dependent variable is the estimated DA–RUM dispersion parameter $\hat{\sigma}$. Column (1) restricts the sample to subjects with exact GARP consistency. Column (2) reports the baseline consistent-sample specification from Table 5. Columns (3) and (4) restrict the sample to subjects with fGARP and EU scores above their respective 95th-percentile thresholds. All specifications include study fixed effects. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

H DA–RUM Implied Risk Aversion and Deterministic Recovery

Table A10 reports the distribution of the local risk-aversion index $r(1)$ implied by the recovered DA–RUM parameters $(\hat{\rho}_i, \hat{a}_i)$. Following Choi et al. (2007), the index is approximated by

$$r(1) \approx \frac{\hat{a}-1}{\hat{a}+1} + \hat{\rho}\frac{2\hat{a}}{(\hat{a}+1)^2}.$$

The measure combines CRRA curvature and disappointment aversion into a scalar summary of local risk attitudes around certainty.

Three features are noteworthy. First, the DA–RUM implied estimates are substantially less extreme than the deterministic benchmark reported by Choi et al. (2007). Their mean deterministic $r(1)$ estimate is 0.919, whereas the mean DA–RUM implied $r(1)$ in Study 1 is 0.355. The difference is consistent with the stochastic model assigning part of the residual variation in observed choices to $\hat{\sigma}$, rather than requiring the preference parameters alone to rationalise every deviation from the deterministic optimum.

Second, the DA–RUM produces comparatively compressed upper tails. The 95th percentile of $r(1)$ is 0.624 in Study 1, 0.950 in Study 2, 0.676 in Study 3, and 0.674 in Study 4. The estimates therefore remain below one at the 95th percentile in all four samples. This pattern is consistent with structural dispersion limiting the extent to which residual behavioural variation appears as extreme recovered risk attitudes.

Third, the central tendency of recovered risk attitudes is relatively stable across the four studies. Median $r(1)$ ranges from 0.314 in Study 1 to 0.356 in Study 2, while the corresponding means range from 0.355 to 0.435. The upper tails vary more than the medians, with Study 2 displaying the greatest dispersion. The common central tendency suggests that the DA–RUM recovers broadly comparable local risk attitudes across environments, while allowing residual choice variation to differ through $\hat{\sigma}$.

The comparison with the deterministic estimates should be interpreted with appropriate caution because the recovery criteria and estimation samples are not identical. Nevertheless, the pattern is consistent with the central interpretation developed in the paper: modelling structural dispersion reduces the need for the deterministic preference parameters to absorb all observed choice variation. It does not eliminate the consequences of misspecification, but it yields a more compressed and cross-study stable distribution of recovered local risk attitudes.

I Structural Dispersion across Recovered Risk Attitudes

Table A10: Distribution of DA–RUM Implied Local Risk Aversion $r(1)$

	Study 1	Study 2	Study 3	Study 4
Mean	0.355	0.435	0.365	0.355
Std. Dev.	0.161	0.241	0.168	0.144
$p5$	0.181	0.160	0.152	0.126
$p25$	0.257	0.261	0.248	0.266
$p50$	0.314	0.356	0.349	0.338
$p75$	0.405	0.553	0.436	0.431
$p95$	0.624	0.950	0.676	0.674
Observations	24	644	105	133

Notes. The table reports the distribution of the DA–RUM implied local risk-aversion index $r(1)$, calculated using the approximation in Choi et al. (2007). The sample is restricted to observations with converged DA–RUM estimates, $1 \leq \hat{a} < 15$, and nonmissing $\hat{\rho}$ and $\hat{a}$. Estimates at the expected-utility boundary, $\hat{a} = 1$, are retained. At this boundary, the index reduces to $\hat{\rho}/2$.

Table A11: Structural Dispersion across Quartiles of Recovered Risk Aversion

$r(1)$ quartile	N	Mean $r(1)$	Mean $\hat{\sigma}$	Median $\hat{\sigma}$	Share high-$\hat{\sigma}$	Mean $\hat{\rho}$	Mean $\hat{a}$
Q1 (lowest)	227	0.194	0.281	0.246	0.436	0.271	1.139
Q2	226	0.304	0.241	0.228	0.363	0.428	1.224
Q3	227	0.422	0.155	0.123	0.141	0.470	1.571
Q4 (highest)	226	0.733	0.074	0.015	0.058	0.658	4.000

Notes. The table reports averages by quartile of implied DA–RUM local risk aversion $r(1)$. The sample includes observations with converged DA–RUM estimates, $\hat{a} < 15$, and nonmissing $\hat{\rho}$ and $\hat{a}$. Estimates at the expected-utility boundary, $\hat{a} = 1$, are retained. High-$\hat{\sigma}$ denotes membership in the top quartile of the estimated structural dispersion distribution.

Table A12: Recovered Risk Aversion and Structural Dispersion

	(1) $\hat{\sigma}_i$	(2) $\hat{\sigma}_i$
$r(1)$ quartile 2	-0.0419**	
	(0.017)	
$r(1)$ quartile 3	-0.123***	
	(0.017)	
$r(1)$ quartile 4	-0.216***	
	(0.017)	
Centred $r(1)$		-0.443***
		(0.053)
Centred $r(1)^2$		0.180
		(0.121)
Study fixed effects	Yes	Yes
Observations	906	906
R^2	0.246	0.277

Notes. Robust standard errors are reported in parentheses. The omitted category in Column (1) is the lowest quartile of the implied DA–RUM local risk-aversion index $r(1)$. Column (2) includes $r(1)$, centred at its sample mean, and its square. All specifications include study fixed effects. The sample is restricted to observations with converged DA–RUM estimates, $1 \leq \hat{a} < 15$, and nonmissing $\hat{\rho}$ and $\hat{a}$. Estimates at the expected-utility boundary, $\hat{a} = 1$, are retained. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.