Munich Personal RePEc Archive

REDRUM: Robust Estimation and Design for the Random Utility Model

Conte, Anna and Hey, John D.

Sapienza University of Rome, University of York

18 May 2026

Online at https://mpra.ub.uni-muenchen.de/129162/
MPRA Paper No. 129162, posted 05 Jun 2026 15:06 UTC

REDRUM: Robust Estimation and Design for the Random Utility Model

Anna Conte[1]

John D. Hey [2]

[1]Department of Statistical Sciences, Sapienza University of Rome, Italy
[2]University of York, UK

anna.conte@uniroma1.it
john.hey@york.ac.uk

Abstract

Non-monotonicity of utility differences in CRRA-based Random Utility Models has been taken to suggest possible distortions in the estimation of risk preferences. We show that this concern is real only under restrictive assumptions on the stochastic component of the model. Under common scaling, changes in utility differences are absorbed by the precision parameter when this is estimated jointly with risk aversion, so that the likelihood in the risk aversion parameter is invariant. We then show that when different subsets of tasks are represented at different effective scales, distortions may arise if a common error scale is imposed, whereas allowing for parsimonious heteroskedasticity restores stable estimation. Simulations based on standard lottery tasks illustrate these results and clarify that the empirical relevance of non-monotonicity depends not on scaling *per se*, but on how the stochastic component is specified.

Keywords: random utility model; risk aversion; CRRA; experimental design; monotonicity; scaling

JEL Classification: C91; D81; C13; C18; C52

1. Introduction

In laboratory experiments, subjects rarely make choices over lotteries expressed in euros or dollars. Instead, outcomes are typically framed in artificial units such as tokens or Experimental Currency Units (ECUs), which are later converted into real payments. In what follows, we refer to such transformations as *scaling*: more generally, any multiplicative transformation of outcomes by a positive constant. The conversion of experimental currency units into monetary payoffs is one natural example of such a transformation.

This raises a simple but largely overlooked question: When estimating risk preferences from such data, should the researcher use outcomes expressed in tokens, or their conversion into monetary units?

At first glance, the answer seems trivial. Under standard expected utility theory with CRRA preferences, scaling all outcomes by a positive constant should not affect preferences. Therefore, whether one uses tokens or euros should be irrelevant.

However, it has been noted by Apesteguia and Ballester (2018) that the difference in expected utilities between lotteries may be non-monotonic in the risk aversion parameter. This non-monotonicity could, in principle, affect the shape of the likelihood function and complicate estimation. This creates an apparent paradox: the units in which outcomes are expressed may determine whether the likelihood is monotonic or not.

In this paper, we show that this concern is largely misplaced. While scaling outcomes can change the shape of the utility difference, it need not affect the estimation of the risk aversion parameter in standard Random Utility Models (RUMs), provided the stochastic component is specified consistently with the scale at which utility differences enter the model.

Apesteguia and Ballester (2018) argue that non-monotonicity in the utility difference may complicate the identification of risk preferences, and may lead to biased estimates of the risk aversion parameter. Our results suggest that the practical relevance of this concern depends crucially on the stochastic specification of the model. Under common scaling, the resulting change in utility differences is absorbed by the precision parameter when this is estimated jointly with risk aversion. More generally, when tasks differ in their effective scale, distortions arise only if the model imposes a common error scale where the stochastic component should be allowed to vary.

Our contribution is threefold:

- We show analytically that scaling affects monotonicity but not indifference points.
- We demonstrate that, in a RUM framework, scaling leaves the likelihood in the risk aversion parameter unchanged once error is properly accounted for (by estimating the error parameter in addition to the risk aversion parameter.)
- Using simulations based on standard experimental designs, we show that estimated risk preferences are invariant to scaling.

What matters in practice. We fully agree that non-monotonicity of the utility difference is a genuine theoretical feature of the model. Our point is different: in experimental applications, this feature is largely irrelevant. Non-monotonicity is a global property, but experiments are designed locally, around empirically plausible values of the risk aversion parameter. In this sense, the issue is not one of identification, but of estimation.

Experimental design does not resolve non-monotonicity, nor is it required to do so. Rather, it ensures that the chosen tasks are informative in the relevant range of the parameter. Within this range, the behaviour of the utility difference is well behaved for the purposes of estimation, and any scaling effects are absorbed by the error parameter.

What matters in practice is therefore not whether the model is globally monotone, but whether the experimental design yields informative variation in the region of interest.

> Non-monotonicity is a global theoretical property; estimation in experiments is a local problem.

Scaling *per se* is not problematic: what matters is whether it is uniform. Uniform scaling is absorbed by the stochastic component, while non-uniform scaling distorts the mapping between utility differences and choice probabilities.

The remainder of the paper is organised as follows. Section 2 introduces the theoretical framework and discusses the relationship between utility scaling and stochastic specification in the CRRA Random Utility Model. Section 3 examines the distinction between scale normalisation and scale identification. Section 4 presents the simulation evidence. Section 5 discusses the implications of the results for experimental design and estimation. Section 6 concludes.

2. Scaling, Non-Monotonicity, and the Random Utility Model

We begin by introducing the object at the centre of the analysis: the difference in expected utilities between two lotteries. For any pair of lotteries x and y, define

$$\Delta(\omega) = E[u(Z_x)] - E[u(Z_y)],$$

where $u(\cdot)$ is the CRRA utility function and ω is the coefficient of relative risk aversion.

This object plays a dual role. It determines indifference points, through the condition $\Delta(\omega^*) = 0$, and it also governs choice probabilities in the Random Utility Model (RUM).

Scaling. We define *scaling* as a multiplicative transformation of all outcomes by a positive constant $k > 0$. For any lottery Z, scaling implies

$$Z \mapsto kZ,$$

and hence, under CRRA preferences,

$$\Delta_k(\omega) = k^{1-\omega}\Delta(\omega).$$

Monotonicity and non-monotonicity. We say that the utility difference $\Delta(\omega)$ is monotone if it is either non-increasing or non-decreasing in ω. Non-monotonicity arises when $\Delta(\omega)$ changes direction, that is, when there exist values $\omega_1 < \omega_2 < \omega_3$ such that

$$\Delta(\omega_2) > \Delta(\omega_1) \quad \text{and} \quad \Delta(\omega_2) > \Delta(\omega_3),$$

or the reverse inequalities.

Non-monotonicity implies that the mapping from ω to utility differences is not one-to-one, so that multiple values of ω may generate the same ordering between lotteries. This is the phenomenon highlighted in the literature and at the origin of the concerns about identification.

Proposition 1 (Non-monotonicity and choice probabilities)**.** *Under CRRA preferences, the mapping $\omega \mapsto \Delta(\omega)$ is not, in general, monotone. As a result, depending on the stochastic specification, the probability of choosing a given lottery need not vary monotonically with ω.*

Full proofs and characterisations are provided in Appendix Sections A.1–A.4.

Scaling outcomes has two distinct effects. These properties are derived formally in Appendix Sections A.2–A.4. First, it does not affect indifference points: the value of ω at which a decision maker is indifferent between two lotteries remains unchanged. Second, it does affect the shape of $\Delta(\omega)$, and in particular its monotonicity properties.

These two facts may appear to create a tension. If scaling changes the shape of the utility difference, it might also affect the estimation of preferences. The key insight of this paper is that this is not the case once the stochastic component of the model is properly specified.

In a RUM framework, scaling changes the magnitude of utility differences, but this change can be absorbed by the error (precision) parameter when it is estimated jointly with ω. As a result, the likelihood function in ω is invariant to scaling.

Reading guide. The discussion below is structured at two levels. The main text presents the argument in a conceptual and intuitive way. Short technical notes provide intermediate analytical steps for readers interested in the underlying mechanisms, while full formal proofs are reported in the Appendix. These technical notes can be skipped without loss of continuity.

Technical note 1: Scaling and indifference points

Under CRRA preferences, scaling all outcomes by a factor $k > 0$ implies

$$\Delta_k(\omega) = k^{1-\omega}\Delta(\omega).$$

Since $k^{1-\omega} > 0$, it follows that

$$\Delta(\omega^*) = 0 \iff \Delta_k(\omega^*) = 0.$$

Thus, indifference points are invariant to scaling.

Technical note 2: Scaling and non-monotonicity

Differentiating $\Delta(\omega)$ with respect to ω shows that scaling introduces an additional term proportional to $-\ln(k)\Delta(\omega)$. As a result, scaling changes the slope of $\Delta(\omega)$.

Therefore, monotonicity is not invariant to scaling: the same pair of lotteries may generate a monotone or non-monotone utility difference depending on the scale of outcomes.

Technical note 3: Scaling in the Random Utility Model

Consider a Random Utility Model with logistic errors:

$$P(x|\omega, \tau) = \Lambda\big(\tau\Delta(\omega)\big), \quad \Lambda(t) = \frac{1}{1+e^{-t}},$$

where $\tau > 0$ is a precision parameter.

Under scaling,

$$\Delta_k(\omega) = k^{1-\omega}\Delta(\omega),$$

so that choice probabilities become

$$P_k(x|\omega, \tau) = \Lambda\big(\tau k^{1-\omega}\Delta(\omega)\big).$$

If τ is fixed, scaling alters the likelihood function and therefore the estimate of ω.

If τ is estimated jointly with ω, define

$$\tau' = \tau k^{1-\omega}.$$

Then

$$P_k(x|\omega, \tau) = \Lambda\big(\tau'\Delta(\omega)\big),$$

which has the same functional form as the original model.

Thus, under common scaling, the effect of scaling is absorbed by the precision parameter, and the likelihood function in ω is invariant.

Full formal proofs of these results are provided in Appendix Sections A.5–A.7, where the arguments are developed in a systematic and fully general way.

Intuition: the precision parameter as a scaling device.

The key object in the Random Utility Model is not the utility difference $\Delta(\omega)$ alone, but the product

$$\tau \cdot \Delta(\omega),$$

which determines choice probabilities.

This has an important implication. Scaling outcomes changes the magnitude of utility differences, but this change is observationally equivalent to a change in the precision parameter.

Intuitively, τ acts as a *zoom lens* on utility differences. If outcomes are expressed in large units (for example, tokens), $\Delta(\omega)$ becomes large and the estimated precision parameter becomes smaller. If outcomes are expressed in small units (for example, euros), $\Delta(\omega)$ becomes small and the estimated precision parameter becomes larger.

What matters for choice probabilities is therefore not the absolute scale of utility differences, but the extent to which the stochastic component magnifies or compresses them. In this sense, scaling of outcomes and scaling of the stochastic component are observationally equivalent under common scaling.

Proposition 2 (Scale invariance under common scaling)**.** *Under CRRA preferences and logistic errors,*

$$P(x|\omega, \tau) = \Lambda\big(\tau\Delta(\omega)\big).$$

Scaling all outcomes by $k > 0$ is equivalent to the transformation $\tau \mapsto \tau k^{1-\omega}$. Hence, the likelihood function in ω is invariant to scaling.

Discussion. Under common scaling, changes in the magnitude of outcomes are observationally equivalent to changes in the scale of the stochastic component. As a consequence, scaling does not alter the likelihood function in ω when the precision parameter is freely estimated. This implies that non-monotonicity, by itself, need not generate estimation problems.

The corresponding formal derivation is reported in Appendix Sections A.5–A.7.

Proposition 3 (Monotone representation under common scaling)**.** *Consider a set of decision tasks indexed by $t = 1, \ldots, T$, with utility differences*

$$\Delta_t(\omega).$$

Suppose there exists a common scaling factor $k > 0$ such that, for every task t, the scaled utility difference

$$\Delta_{t,k}(\omega) = k^{1-\omega}\Delta_t(\omega)$$

is monotone in ω over the parameter region of interest. Then the original model is observationally equivalent to a monotone representation of the same model. In particular, estimation of ω is unaffected by using the monotone scaled representation, provided the precision parameter τ is freely estimated.

Proof sketch. Under CRRA preferences, common scaling by $k > 0$ transforms each utility difference as

$$\Delta_t(\omega) \mapsto \Delta_{t,k}(\omega) = k^{1-\omega}\Delta_t(\omega).$$

In the logit RUM, choice probabilities are

$$P_t(\omega, \tau) = \Lambda\big(\tau\Delta_t(\omega)\big).$$

After scaling, they become

$$P_{t,k}(\omega, \tau) = \Lambda\big(\tau k^{1-\omega}\Delta_t(\omega)\big).$$

Define

$$\tau_k(\omega) = \tau k^{1-\omega}.$$

Then

$$P_{t,k}(\omega, \tau) = \Lambda\big(\tau_k(\omega)\Delta_t(\omega)\big),$$

which has the same functional form as the original model. Since τ is freely estimated, the scaling transformation is absorbed by a reparametrisation of the precision parameter.

Therefore, if a common scaling factor produces a representation in which all $\Delta_{t,k}(\omega)$ are monotone over the relevant range, this monotone representation is observationally equivalent to the original one. The likelihood in ω is unchanged, and estimation of the risk aversion parameter is unaffected. □

A more general characterisation of the scaling factor required to induce monotonicity is given in Appendix Section A.4.1.

2.1. Global vs. Local Scaling and Experimental Design

The invariance result established above relies on a common multiplicative transformation of all outcomes. When all payoffs are scaled by the same factor, the resulting change in utility differences can be absorbed by a single precision parameter, leaving the likelihood function in ω unchanged.

This property, however, does not extend to situations in which different subsets of tasks are scaled by different factors. The formal distinction between common and non-uniform scaling is summarised in Appendix Sections A.8–A.9. Suppose that two groups of decision tasks are scaled by k_1 and k_2, respectively. Then the utility differences are transformed as

$$\Delta_t(\omega) \mapsto k_t^{1-\omega}\Delta_t(\omega),$$

where $k_t \in \{k_1, k_2\}$. In this case, the induced task-specific multipliers cannot, in general, be absorbed by a single common precision parameter. As a result, scale invariance fails unless $k_1 = k_2$ (or in the special case $\omega = 1$).

This observation helps clarify the nature of the non-monotonicity problem. The non-monotonicity result is global, in the sense that it applies to arbitrary collections of decision tasks without restrictions on their scale or composition. In such a general setting, inconsistencies induced by scaling may indeed arise.

By contrast, experimental applications are inherently local. In laboratory settings, decision tasks are designed by the researcher and are typically expressed in a common unit (tokens or experimental currency units) with a fixed conversion rate into monetary payoffs. Budget constraints and practical considerations naturally restrict the range of outcomes that can be implemented, and hence the range of risk attitude that can be meaningfully explored.

As a consequence, experimental data sets correspond to collections of tasks that can be designed in a way to share a common scale. Within such environments, the invariance result applies directly: any global rescaling of outcomes is absorbed by the precision parameter and does not affect the estimation of ω.

More broadly, this highlights the role of experimental design. Researchers do not face arbitrary collections of lotteries, but can construct decision tasks strategically so as to be informative over a plausible range of the parameter. In this context, the relevant issue is not whether the model is globally monotone, but whether the chosen tasks provide sufficient variation in the region of interest. Non-monotonicity remains a theoretical possibility, but it does not undermine estimation in the experimental settings for which the model is intended.

3. Scale Normalisation vs. Scale Identification

A well-known response to the lack of scale invariance in CRRA-based random utility models is to normalise outcomes within each choice problem. In particular, Wilcox (2011) proposes specifying utility relative to task-specific reference points, typically based on the range of outcomes within each choice set. This leads to a normalised representation of the form

$$\frac{u(x) - u(x_{\min})}{u(x_{\max}) - u(x_{\min})},$$

where $u(\cdot)$ is the CRRA utility function, which rescales utility differences so that they are invariant to proportional transformations of all outcomes within the task.

This normalisation ensures that proportional rescaling of all outcomes leaves utility differences unchanged, since both the numerator and denominator scale in the same way. As a result, the deterministic component of the latent utility index becomes invariant to outcome scaling.

While effective in eliminating scale dependence, this approach comes at a cost. First, the transformation is not affine and therefore alters the underlying preference representation. In particular, the curvature of utility—and hence the implied degree of risk aversion—depends on the distribution of outcomes within the choice set. This introduces a form of contextual dependence that is absent in standard expected utility models.

Secondly, the normalisation is innocuous only in simple environments, such as binary-outcome lotteries, where scaling relative to the range preserves the relevant structure of the problem. In more general settings with multiple outcomes, normalising by a single range compresses all outcomes relative to the extremes, thereby altering the relative spacing of intermediate outcomes. Since CRRA utility is nonlinear, this transformation affects curvature in a non-uniform way and is not equivalent to an affine transformation of utility. As a result, the induced utility differences no longer correspond to those generated by the original preference specification.

This perspective also helps clarify the interpretation of the results in Apesteguia and Ballester (2018). While their analysis does not explicitly discuss the scale of the stochastic component, it can be understood as operating under an implicit normalisation in which the error scale is fixed and common across tasks. Under this normalisation, differences in the scaling of outcomes across decision problems are not absorbed by the stochastic component, and therefore translate into differences in the magnitude of utility differences that affect estimation.

In this sense, their findings do not reflect a general failure of scale invariance, but rather the consequences of imposing a fixed error scale in the presence of heterogeneous outcome scaling.

By contrast, our approach preserves the original utility specification and instead relaxes the normalisation imposed on the stochastic component of the model. Under CRRA preferences, scaling all outcomes by a factor $k > 0$ multiplies utility differences by $k^{1-\omega}$. In a standard RUM with fixed error variance, this alters choice probabilities and hence the likelihood function. However, if the scale parameter of the error term is freely estimated, the effect of outcome scaling can be absorbed by a corresponding transformation of the noise scale.

Formally, scaling outcomes induces the transformation $\Delta(\omega) \mapsto k^{1-\omega}\Delta(\omega)$, which is observationally equivalent to a rescaling of the precision parameter. As a consequence, the likelihood function in ω is invariant to outcome scaling, and the estimation of risk preferences is unaffected.

Wilcox normalises utility differences directly, whereas we show that this is unnecessary once the stochastic component is properly specified. The key distinction is therefore conceptual. Outcome normalisation restores invariance by modifying the utility representation, effectively embedding contextual information into preferences. Our approach, instead, restores invariance within the standard framework by recognising that the scale of the stochastic component is not identified independently of utility differences.

In this sense, scale invariance is not a property that must be imposed on preferences, but one that emerges naturally once the model is correctly specified. This perspective clarifies that the apparent scale dependence of estimated risk aversion in standard CRRA-RUM models is not a fundamental feature of preferences, but rather a consequence of an arbitrary normalisation of the error term.

3.1. Non-Uniform Scaling Across Tasks

The invariance result established above relies on a common multiplicative transformation of all outcomes. In experimental applications, however, different subsets of tasks may involve systematically

Table 1: Alternative approaches to scale dependence in CRRA-RUM models

Approach	Mechanism	Implication for utility	Invariance in ω
Apesteguia & Ballester (2018)	Fixed error scale	Utility unchanged	No
Wilcox (2011)	Normalisation of utility differences	Utility representation modified	Yes
This work	Estimation of error scale	Utility unchanged	Yes

Notes. The table compares three approaches to scale dependence in random utility models with CRRA preferences. Apesteguia and (2018) highlight the lack of invariance when the error scale is fixed. Wilcox (2011) restores invariance by normalising utility differences using task-specific reference points, thereby modifying the utility representation. Our approach restores invariance by allowing the scale of the stochastic component to be freely estimated, without altering preferences.

different payoff magnitudes.

Formally, let tasks be indexed by t, and suppose that each task is associated with a scaling factor k_t. Under CRRA preferences, utility differences become

$$\Delta_t(\omega) \mapsto k_t^{1-\omega} \Delta_t(\omega).$$

Deterministic vs. stochastic scaling. The transformation above operates on the *deterministic* component of the model, namely the utility difference $\Delta_t(\omega)$. In contrast, the stochastic component enters through the precision parameter τ, which governs the dispersion of the error term.

A key observation is that these two components are observationally equivalent. Since choice probabilities depend on the product $\tau \cdot \Delta_t(\omega)$, a multiplicative transformation of $\Delta_t(\omega)$ is equivalent to an inverse transformation of τ. Thus, scaling of outcomes (deterministic) and heteroskedasticity in the error term (stochastic) are not separately identified.

When scaling is common across tasks, this equivalence allows the transformation to be absorbed by a single precision parameter. However, when k_t varies across tasks, the induced multipliers $k_t^{1-\omega}$ differ across observations. In general, these differences cannot be absorbed by a single common precision parameter.

Heterogeneous scaling and heteroskedasticity. A natural extension of the model allows the stochastic component to vary across tasks. Let tasks be partitioned into groups indexed by j, each associated with a precision parameter τ_j. Choice probabilities can then be written as

$$P_t = \Lambda\big(\tau_j \Delta_t(\omega)\big),$$

where j identifies the group to which task t belongs.

More generally, one can specify a parametric form for heteroskedasticity:

$$\tau_t = \exp(z_t'\gamma),$$

where z_t is a vector of observable characteristics of the task. This formulation is analogous to scale functions used in heteroskedastic Tobit and discrete choice models.

Choice of scale shifters. In experimental settings, heterogeneity in τ_t can be related to observable

characteristics of the tasks, such as the magnitude of payoffs or the dispersion of outcomes. More generally, any feature that affects the scale of utility differences may serve as a determinant of the stochastic component. This provides a simple and flexible way to model heterogeneity in the effective scale of utility across tasks.

Graphical intuition and scaling choices. In practice, researchers often inspect the distribution of utility differences across tasks to assess scale heterogeneity. A simple graphical diagnostic is to plot $\Delta_t(\omega)$ across tasks for plausible values of ω, highlighting differences in magnitude. Such analysis can guide the choice of a common scaling factor (e.g. dividing by a large constant such as 1500) to reduce dispersion. We illustrate this approach in the Appendix.

Implications for estimation. This discussion highlights that the issue is not scaling *per se*, but the interaction between payoff scale and the stochastic specification of the model.

When scaling is uniform, its effect is absorbed by the precision parameter and estimation of ω is unaffected. When scaling varies across tasks, imposing a common error scale forces heterogeneous utility differences to be reconciled through a single noise parameter. The model then compensates by distorting the estimate of ω.

Allowing for heterogeneity in the error scale provides a natural way to absorb such differences and restore a correct mapping between utility differences and choice probabilities. The equivalence between heterogeneous scaling and heteroskedastic precision parameters is formalised in Appendix Section A.9.

Whether such heterogeneity matters in practice is ultimately an empirical question, since the researcher does not observe the data-generating process and may not be able to determine *a priori* whether a common error scale is adequate. In many applications, however, simple graphical diagnostics can provide useful guidance. For instance, plotting $\Delta_t(\omega)$ across tasks for plausible values of ω can reveal whether utility differences are of comparable magnitude or exhibit systematic scale heterogeneity.

Scaling is innocuous only when it is uniform. Once scaling varies across observations, it becomes a source of misspecification that the model resolves by distorting preferences. This is summarised in the following diagram.

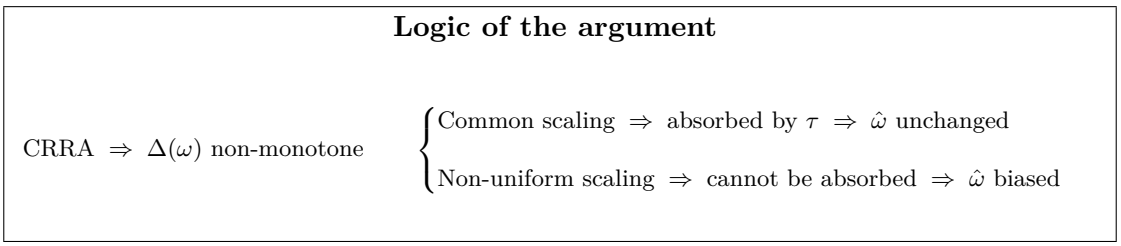

Figure 1: Scaling affects estimation only when it varies across tasks.

4. Simulation Evidence

The purpose of this section is to illustrate how scaling interacts with the stochastic specification of the Random Utility Model. We consider three exercises: (i) common scaling of all tasks, (ii) common scaling with fixed error variance, and (iii) non-uniform scaling across tasks.

4.1. Data Generating Process

We consider 40 binary lottery tasks, reported in Appendix Table 3. The tasks are based on the experimental design of Andersen et al. (2008), adapted to the purposes of the present analysis.

Choices are generated independently from a CRRA Random Utility Model with logistic errors:

$$P_t = \Lambda\big(\tau_0 \Delta_t(\omega_0)\big).$$

The true parameters are:

$$\omega_0 = 0.5, \quad \tau_0 = 2.$$

For each replication, we simulate $N = 5{,}000$ independent choices. Tasks are assigned randomly across observations, with each task appearing $N/40 = 125$ times on average, so that the simulated dataset is approximately balanced across tasks.

We repeat the experiment 1,000 times and report average estimates across replications.

4.2. Simulation Design

We consider two distinct experimental designs that differ only in how payoff scaling is introduced at the estimation stage.

Experiment 1: Common scaling. All payoffs are uniformly rescaled by dividing outcomes by a constant factor $k = 1500$:

$$(x, y) \mapsto \left(\frac{x}{1500}, \frac{y}{1500}\right).$$

Choices are always generated from the original (unscaled) tasks.

Experiment 2: Non-uniform scaling. Tasks are partitioned into two groups. For tasks in group 1, payoffs remain unchanged. For tasks in group 2, payoffs are divided by 100:

$$(x, y) \mapsto (x, y) \quad \text{or} \quad \left(\frac{x}{100}, \frac{y}{100}\right).$$

Again, choices are generated from the original homogeneous DGP, and scaling is introduced only at the estimation stage.

Across both experiments, we estimate alternative specifications of the Random Utility Model that differ in the treatment of the precision parameter.

4.3. Simulation Results

Table 2 reports the Monte Carlo results. The evidence is organised by panels corresponding to the different specifications of the model.

Table 2: Monte Carlo results

Panel	Specification	$\hat{\omega}$	Bias($\hat{\omega}$)	s.e.($\hat{\omega}$)	$\hat{\tau}$
DGP: $\omega_0 = 0.5$, $\tau_0 = 2$, 1,000 replications, 5,000 observations					
A	Original scale, τ estimated	0.499940	-0.000060	0.002198	2.021
A	Payoffs divided by 1500, τ estimated	0.499940	-0.000060	0.002198	78.297
B	Original scale, τ fixed	0.499837	-0.000163	0.002142	fixed at 2
B	Payoffs divided by 1500, τ fixed	0.667966	0.167966	0.018149	fixed at 2
C	Original representation, single τ	0.499944	-0.000056	0.002193	2.025
C	Mixed representation, single τ	0.477219	-0.022781	0.001132	10.703
C	Mixed representation, (τ_1, τ_2)	0.500001	0.000001	0.002543	(2.058, 20.489)

Notes. The table reports Monte Carlo averages. In Panels A and B, common scaling means that all payoffs are divided by 1500 at the estimation stage. In Panel C, mixed representation means that tasks 1–20 are left at their original scale, while tasks 21–40 are divided by 100 at the estimation stage. Bias is computed relative to $\omega_0 = 0.5$. The fixed-τ specification fixes τ at its true DGP value, $\tau_0 = 2$.

Common scaling (Panel A and Panel B). When the precision parameter is estimated (Panel A), scaling has no effect on the estimated risk aversion parameter. The mean estimate remains $\hat{\omega} \approx 0.500$, and its standard error is unchanged. By contrast, the estimated precision parameter adjusts substantially, from approximately 2 to about 78, consistent with the theoretical scaling relationship.

When the precision parameter is fixed (Panel B), scaling induces a large distortion. The estimate shifts to $\hat{\omega} \approx 0.668$, generating a bias of approximately 0.17. This reflects the fact that the model cannot absorb the change in the magnitude of utility differences through the stochastic component.

Non-uniform scaling (Panel C). In the baseline case with a common scale (Panel C, first row), the estimate is unbiased, $\hat{\omega} \approx 0.500$.

When tasks are combined at different payoff scales and a single precision parameter is imposed (Panel C, second row), the estimate becomes distorted, $\hat{\omega} \approx 0.477$, corresponding to a bias of about -0.023.

Allowing for group-specific precision parameters (Panel C, third row) restores the estimate. The mean returns to $\hat{\omega} \approx 0.500$, while the estimated precision parameters differ across groups, with $\hat{\tau}_1 \approx 2$ and $\hat{\tau}_2 \approx 20$.[1]

[1]The magnitude of the change in the estimated precision parameter closely matches the theoretical prediction. For instance, with a scaling factor of $k = 1/1500$ and $\omega = 0.5$, the theory implies $\tau' = \tau \cdot 1500^{1-\omega} \approx 2 \cdot 38.73 = 77.46$, which is very close to the simulated value of 78.30. A similar correspondence is observed in the non-uniform scaling case, where $k = 1/100$ implies $\tau' \approx 2 \cdot 10 = 20$, compared to the estimated 20.49.

4.4. Interpretation

The simulation results closely match the theoretical predictions. In particular, scaling changes the magnitude of utility differences, which is observationally equivalent to a change in the scale of the stochastic component.

When this scale is estimated, the effect is absorbed by the precision parameter and leaves ω unchanged. When it is constrained—either by fixing τ or imposing a common τ across heterogeneous tasks—the model compensates by distorting ω.

Allowing for heterogeneity in the error scale restores invariance.

As a robustness check, we replicate the Monte Carlo exercises with larger samples, using $N = 10{,}000$ and $N = 25{,}000$ observations per replication. The results, reported in Appendix B.3 Table 4, confirm the conclusions from the baseline simulation. Under common scaling, the estimate of ω is unchanged when τ is freely estimated, while fixing τ generates the same systematic upward distortion. Under non-uniform scaling, imposing a single precision parameter biases $\hat{\omega}$ downward, whereas allowing for group-specific precision parameters restores the estimate. Increasing the sample size reduces sampling variability but does not alter the qualitative pattern.

5. Discussion

The analysis clarifies the role of non-monotonicity in empirical work.

Non-monotonicity is a genuine theoretical feature of the CRRA specification, and scaling can affect the shape of the utility difference in ways that alter its monotonicity properties. However, this does not imply that the estimation of risk preferences is necessarily unstable or biased in experimentally relevant settings.

What matters empirically is the interaction between utility differences and the stochastic component of the model. Under common scaling, changes in the magnitude of utility differences are absorbed by the precision parameter when this is estimated jointly with risk aversion. More generally, when tasks differ in their effective scale, the practical relevance of non-monotonicity depends on whether the stochastic specification is flexible enough to accommodate such heterogeneity.

This perspective helps clarify the relation between our results and those of Apesteguia and Ballester. Their analysis is conducted in a general setting and highlights a genuine global property of the model. Our point is not that this property disappears, but that its empirical consequences depend on the structure of the tasks and on how the stochastic component is specified. In standard experimental environments, where tasks are often expressed on a common or similar scale, and where the error scale is estimated rather than fixed, the resulting variation in utility differences may have limited practical impact on the estimation of risk preferences.

The broader methodological implication is that apparent scale dependence should not automatically be interpreted as a failure of the preference specification. In many cases, it reflects instead a mismatch between the effective scale of utility differences and the imposed scale of the stochastic

component. From this perspective, allowing for parsimonious heteroskedasticity is not a correction to the model, but a safeguard against potential misspecification.

Non-monotonicity is therefore best understood as a theoretical property whose empirical relevance is context-dependent. In experimental applications, the central question is not whether the model is globally monotone, but whether the chosen tasks are informative in the relevant range of the parameter and whether the stochastic specification is consistent with the scale at which those tasks are effectively evaluated.

Uniform scaling is a reparametrisation; non-uniform scaling is a misspecification.

6. Conclusion

This paper revisits the role of outcome scaling in random utility models with CRRA preferences. We show that scaling *per se* is not problematic: under common scaling, changes in the magnitude of utility differences are observationally equivalent to changes in the scale of the stochastic component, leaving preference estimates unaffected when the precision parameter is freely estimated.

The problem arises when scaling is heterogeneous across tasks. In this case, a model with a common error scale cannot accommodate differences in the magnitude of utility differences, and the resulting misspecification is absorbed by distortions in the estimated preference parameter. Allowing for heterogeneity in the stochastic component restores identification without modifying the utility representation.

More broadly, our results clarify that scale dependence in CRRA-RUM models is not a feature of preferences, but a consequence of how the stochastic component is specified. This perspective provides a unified interpretation of existing approaches and suggests that modelling the error scale, rather than normalising outcomes, is the natural way to restore invariance.

Several questions remain open. Extending the analysis to alternative preference specifications such as CARA, characterising conditions for monotonicity of utility differences across tasks, and developing guidance for modelling heteroskedastic errors in experimental data are promising directions for future research. More generally, understanding how experimental design interacts with scaling and identification remains an important avenue for further work.

References

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting Risk and Time Preferences. *Econometrica*, 76(3): 583–618.

Apesteguia, J., and Ballester, M. A. (2018). Monotone Stochastic Choice Models: The Case of Risk and Time Preferences. *Journal of Political Economy*, 126(1): 74–106.

Wilcox, N. (2011). Stochastically more risk averse: A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162(1): 89–104.

Appendix A: Formal Results

This appendix collects the formal results underlying the arguments in the main text. Each subsection isolates a specific mechanism—scaling, monotonicity, or stochastic specification—and shows how it affects identification and estimation.

A.1 Homotheticity of CRRA Utility

We begin by recalling a basic property of CRRA preferences that underlies all subsequent results.

Under CRRA preferences,

$$u(x) = \frac{x^{1-\omega}}{1-\omega}, \quad \omega \neq 1,$$

and for any $k > 0$,

$$u(kx) = k^{1-\omega} u(x).$$

Hence, for any lottery L,

$$E[u(kZ_L)] = k^{1-\omega} E[u(Z_L)].$$

Let

$$\Delta(\omega) = E[u(Z_x)] - E[u(Z_y)].$$

Then

$$\Delta_k(\omega) = k^{1-\omega} \Delta(\omega).$$

Implication. Scaling multiplies utility differences by a positive factor, leaving preferences unchanged.

A.2 Scale Invariance of Indifference Points

We now show formally that scaling does not affect the identification of preferences.

Let ω^* satisfy

$$\Delta(\omega^*) = 0.$$

Then

$$\Delta_k(\omega^*) = k^{1-\omega^*} \Delta(\omega^*) = 0.$$

Implication. Indifference points—and hence preference identification—are invariant to scaling.

A.3 Non-Monotonicity of the Utility Difference

We next characterise the behaviour of $\Delta(\omega)$ with respect to ω. This is important because, in the RUM specification, choice probabilities depend on $\tau \cdot \Delta(\omega)$, so the shape of $\Delta(\omega)$ directly affects identification.

Let

$$\Delta(\omega) = \mathbb{E}[u(Z_x;\omega)] - \mathbb{E}[u(Z_y;\omega)].$$

Then

$$\frac{\partial u(z;\omega)}{\partial \omega} = z^{1-\omega}\left[\frac{1}{(1-\omega)^2} - \frac{\ln z}{1-\omega}\right].$$

Taking expectations,

$$\frac{\partial \Delta(\omega)}{\partial \omega} = \frac{\Delta(\omega)}{1-\omega} - \frac{1}{1-\omega}\mathbb{E}\left[z_x^{1-\omega}\ln z_x - z_y^{1-\omega}\ln z_y\right].$$

The sign of this derivative is not determined in general, as it depends on a weighted difference of logarithmic terms.

Implication. The utility difference $\Delta(\omega)$ need not be monotone in ω. This non-monotonicity is an intrinsic property of the CRRA specification and arises independently of any scaling transformation.

A.4 Effect of Scaling on Monotonicity

We now examine how scaling interacts with the shape of $\Delta(\omega)$.

Under a common scaling of outcomes $z \mapsto kz$, we obtain

$$\Delta_k(\omega) = k^{1-\omega}\Delta(\omega).$$

Differentiating,

$$\frac{\partial \Delta_k}{\partial \omega} = k^{1-\omega}\left[\frac{\partial \Delta}{\partial \omega} - \ln(k)\Delta(\omega)\right].$$

Implication. Scaling affects the slope of $\Delta(\omega)$ through the additional term $-\ln(k)\Delta(\omega)$, and may therefore amplify or attenuate existing non-monotonicities. However, scaling does not alter the location of indifference points, since $\Delta_k(\omega) = 0$ if and only if $\Delta(\omega) = 0$.

In the RUM framework, this transformation has a simple interpretation: since choice probabilities depend on $\tau \cdot \Delta(\omega)$, a common scaling is observationally equivalent to a rescaling of the precision parameter τ. As a result, the effect of scaling on $\Delta(\omega)$ can be absorbed by τ when it is freely estimated.

A.4.1 *Choosing a Scaling Factor to Induce Monotonicity*

We now characterise conditions under which a scaling factor can render the utility difference monotone in the risk aversion parameter over a given interval.

Recall that under scaling by $k > 0$,

$$\Delta_k(\omega) = k^{1-\omega}\Delta(\omega),$$

and

$$\frac{\partial \Delta_k}{\partial \omega} = k^{1-\omega}\left[\frac{\partial \Delta}{\partial \omega} - \ln(k)\Delta(\omega)\right].$$

Proposition 4 (Scaling and Monotonicity)**.** *Let $I \subset \mathbb{R}$ be an interval such that $\Delta(\omega) \neq 0$ for all $\omega \in I$. Define*

$$R(\omega) = \frac{\Delta'(\omega)}{\Delta(\omega)}.$$

Then there exists a scaling factor $k > 0$ such that $\Delta_k(\omega)$ is monotone on I if and only if there exists a constant $c \in \mathbb{R}$ satisfying either

$$c \leq \inf_{\omega \in I} R(\omega) \quad \text{or} \quad c \geq \sup_{\omega \in I} R(\omega).$$

In that case, monotonicity is obtained by choosing $k = e^c$.

Interpretation. Scaling affects monotonicity through the term $-\ln(k)\Delta(\omega)$, which shifts the slope of the utility difference. When the ratio $\Delta'(\omega)/\Delta(\omega)$ is bounded over an interval, an appropriate choice of k can prevent the derivative from changing sign.

Implication. A suitable scaling factor can restore monotonicity locally, even when the original utility difference is non-monotonic. However, such a factor need not exist globally. This reinforces the distinction between global properties of the model and the local nature of experimental design.

A.5 Random Utility Model Setup

We now introduce the stochastic structure used in the paper.

Let

$$D(\omega) = E[U(A)] - E[U(B)].$$

In a logit RUM,

$$P(A \succ B) = \frac{1}{1 + \exp(-\tau D(\omega))},$$

where $\tau > 0$ is a precision parameter.

Under scaling,

$$D(\omega) \mapsto k^{1-\omega} D(\omega).$$

Objective. We study how this transformation affects estimation under different specifications of the error term.

A.6 Case 1: Fixed Precision Parameter

Suppose τ is fixed (e.g. normalised to 1).

After scaling,

$$P_{\text{scaled}}(A \succ B) = \frac{1}{1 + \exp(-k^{1-\omega} D(\omega))}.$$

The likelihood becomes

$$L_{\text{scaled}}(\omega) = \prod_j \left[\frac{1}{1 + \exp(-k^{1-\omega} D_j(\omega_{\text{true}}))} \right]^{y_j} \left[1 - \frac{1}{1 + \exp(-k^{1-\omega} D_j(\omega_{\text{true}}))} \right]^{1-y_j}.$$

Implication. Scaling changes the likelihood function unless $\omega = 1$.

Consistency requires

$$k^{1-\omega^*} D_j(\omega_{\text{true}}) \approx D_j(\omega^*),$$

which implies $\omega^* \neq \omega_{\text{true}}$ when $k \neq 1$.

Conclusion. The estimator of ω is inconsistent.

A.7 Case 2: Estimated Precision Parameter

Suppose τ is estimated jointly with ω.

After scaling,

$$P_{\text{scaled}}(A \succ B) = \frac{1}{1 + \exp(-\tau k^{1-\omega} D(\omega))}.$$

Define

$$\tau' = \tau k^{1-\omega}.$$

Then

$$P_{\text{scaled}}(A \succ B) = \frac{1}{1 + \exp(-\tau' D(\omega))}.$$

Implication. Scaling is equivalent to a reparametrisation of τ.

$$(\omega, \tau) \mapsto (\omega, \tau k^{1-\omega})$$

leaves probabilities unchanged.

Conclusion. The estimator of ω is consistent and invariant to scaling.

A.8 Non-Uniform Scaling Across Tasks

The previous result relies on a common scaling factor. We now consider the case in which different tasks are associated with different effective scales.

Consider now heterogeneous scaling:

$$\Delta_t(\omega) \mapsto k_t^{1-\omega} \Delta_t(\omega).$$

Implication. A single τ cannot absorb task-specific scaling.

Conclusion. The standard RUM is misspecified.

A.9 Heteroskedastic Extension

Allow group-specific precision parameters:

$$P_t = \Lambda(\tau_j \Delta_t(\omega)).$$

Implication. Heterogeneous scaling is observationally equivalent to heteroskedasticity.

Conclusion. Allowing τ_j provides a parsimonious way of absorbing task-specific scale heterogeneity in the stochastic component.

A.10 Summary

Case	Parameters Estimated	Error parameter	Effect on $\hat{\omega}$
Case 1	ω only	Fixed ($\tau = 1$)	Inconsistent
Case 2	ω and τ	Free	Consistent

A.11 Practical Implications

1. Fixing the error scale induces spurious dependence on outcome units.
2. Estimating the error scale restores invariance and consistent estimation.

Appendix B: Simulation Design and Additional Figures

B.1 Decision Tasks

The lottery tasks used in the simulations are based on the experimental design of Andersen et al. (2008), adapted to the purposes of the present analysis.

Table 3: Decision tasks used in the simulations

Task	x_1	x_2	p_1	y_1	y_2
1	3850	100	0.1	2000	1600
2	3850	100	0.2	2000	1600
3	3850	100	0.3	2000	1600
4	3850	100	0.4	2000	1600
5	3850	100	0.5	2000	1600
6	3850	100	0.6	2000	1600
7	3850	100	0.7	2000	1600
8	3850	100	0.8	2000	1600
9	3850	100	0.9	2000	1600
10	3850	100	1.0	2000	1600
11	4000	500	0.1	2250	1500
12	4000	500	0.2	2250	1500
13	4000	500	0.3	2250	1500
14	4000	500	0.4	2250	1500
15	4000	500	0.5	2250	1500
16	4000	500	0.6	2250	1500
17	4000	500	0.7	2250	1500
18	4000	500	0.8	2250	1500
19	4000	500	0.9	2250	1500
20	4000	500	1.0	2250	1500
21	4000	150	0.1	2000	1750
22	4000	150	0.2	2000	1750
23	4000	150	0.3	2000	1750
24	4000	150	0.4	2000	1750
25	4000	150	0.5	2000	1750
26	4000	150	0.6	2000	1750
27	4000	150	0.7	2000	1750
28	4000	150	0.8	2000	1750
29	4000	150	0.9	2000	1750
30	4000	150	1.0	2000	1750
31	4500	50	0.1	2500	1000
32	4500	50	0.2	2500	1000
33	4500	50	0.3	2500	1000
34	4500	50	0.4	2500	1000
35	4500	50	0.5	2500	1000
36	4500	50	0.6	2500	1000
37	4500	50	0.7	2500	1000
38	4500	50	0.8	2500	1000
39	4500	50	0.9	2500	1000
40	4500	50	1.0	2500	1000

Notes. The decision tasks are based on the lottery design of Andersen et al. (2008), adapted to the purposes of the present analysis. Each task consists of a binary choice between two lotteries, $X = (x_1, p_1;\, x_2, 1-p_1)$ and $Y = (y_1, p_1;\, y_2, 1-p_1)$. Choices in the simulations are generated independently from these tasks. The same probability p_1 applies to the first outcome in both lotteries.

B.2 Representation of Unscaled and Scaled Decision Tasks

This appendix illustrates the behaviour of the individual logit likelihood contributions associated with the lottery tasks used in the simulations.

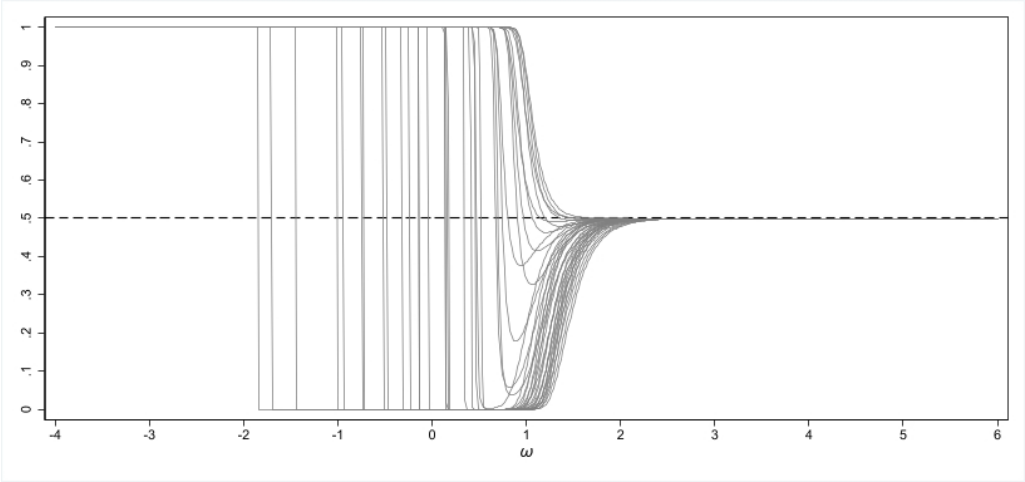

Figure 2: Individual logit likelihood contributions for the original lottery tasks, using $\tau = 2$. Each curve corresponds to one decision task. Dominated tasks are excluded.

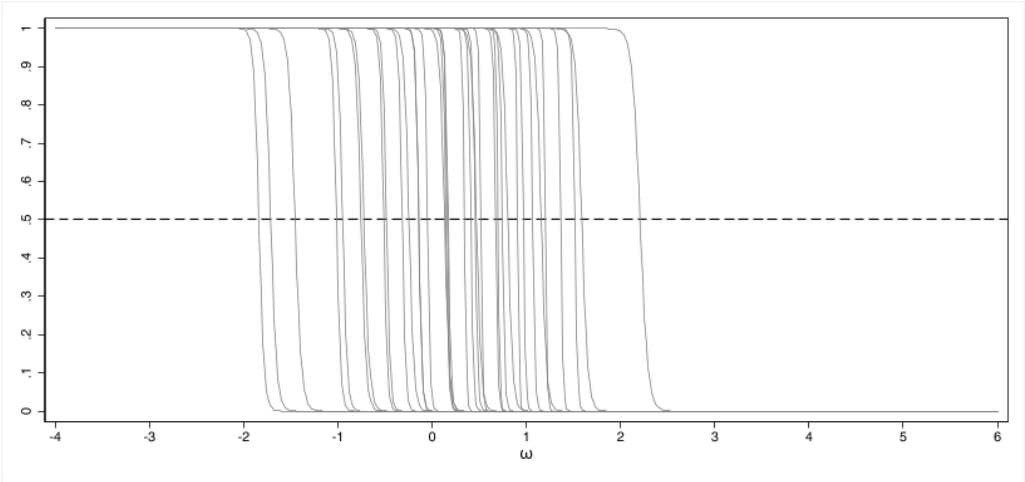

Figure 3: Individual logit likelihood contributions after dividing all outcomes by 1500 and setting $\tau = 77.46$, the theoretical precision implied by common scaling when $\omega = 0.5$. The scaled representation preserves the informational structure of the tasks.

The figures are constructed using the CRRA Random Utility Model with logistic errors. In the unscaled representation, payoffs are expressed in their original units and the precision parameter is set to $\tau = 2$. In the scaled representation, all payoffs are divided by 1500. To compare observationally equivalent representations, the precision parameter is adjusted according to the theoretical scaling relationship:

$$\tau' = \tau \cdot 1500^{1-\omega}.$$

Evaluated at $\omega = 0.5$, this gives

$$\tau' = 2\sqrt{1500} \approx 77.46.$$

Thus, the figures compare two observationally equivalent representations of the same model

rather than imposing the same precision parameter across different outcome scales.

We use 36 tasks, excluding tasks 10, 20, 30, and 40. These excluded tasks involve first-order stochastic dominance and therefore provide no information about the risk aversion parameter ω.

Figure 2 reports the likelihood contributions using the original payoff representation with $\tau = 2$. Figure 3 reports the likelihood contributions after dividing all outcomes by 1500 and setting $\tau' = 77.46$. Under common scaling, the shape of the utility differences changes, but the corresponding change in the precision parameter preserves the informational content of the tasks.

B.3 Robustness to Larger Sample Sizes

The larger-sample simulations confirm the robustness of the baseline findings. In Panel A, common scaling leaves $\hat{\omega}$ unchanged to numerical precision when τ is estimated, while the estimated precision parameter adjusts according to the theoretical scaling relationship. The estimated bias remains numerically negligible for both sample sizes. Small differences in the fifth decimal place reflect Monte Carlo variation rather than a systematic effect of sample size. In Panel B, fixing τ produces a stable upward bias of about 0.168 regardless of sample size. In Panel C, imposing a single precision parameter under non-uniform scaling generates a stable downward bias of about 0.023, whereas allowing for group-specific precision parameters restores $\hat{\omega}$ to its true value. Thus, increasing the sample size reduces standard errors but does not remove bias when the stochastic component is misspecified.

Table 4: Monte Carlo robustness checks: larger samples

Sample size	Specification	$\hat{\omega}$	Bias($\hat{\omega}$)	s.e.($\hat{\omega}$)	$\hat{\tau}$
Panel A: Common scaling, τ estimated					
$N = 10{,}000$	Original scale	0.500005	0.000005	0.001549	2.015
$N = 10{,}000$	Payoffs divided by 1500	0.500005	0.000005	0.001549	78.034
$N = 25{,}000$	Original scale	0.499978	-0.000022	0.000981	2.004
$N = 25{,}000$	Payoffs divided by 1500	0.499978	-0.000022	0.000981	77.638
Panel B: Common scaling, τ fixed					
$N = 10{,}000$	Original scale	0.499922	-0.000078	0.001511	fixed at 2
$N = 10{,}000$	Payoffs divided by 1500	0.668030	0.168030	0.012833	fixed at 2
$N = 25{,}000$	Original scale	0.499980	-0.000020	0.000954	fixed at 2
$N = 25{,}000$	Payoffs divided by 1500	0.668099	0.168099	0.008116	fixed at 2
Panel C: Non-uniform scaling					
$N = 10{,}000$	Original representation, single τ	0.500026	0.000026	0.001548	2.017
$N = 10{,}000$	Mixed representation, single τ	0.477210	-0.022790	0.000795	10.637
$N = 10{,}000$	Mixed representation, (τ_1, τ_2)	0.500065	0.000065	0.001787	(2.033, 20.296)
$N = 25{,}000$	Original representation, single τ	0.499961	-0.000039	0.000981	2.006
$N = 25{,}000$	Mixed representation, single τ	0.477045	-0.022955	0.000489	10.802
$N = 25{,}000$	Mixed representation, (τ_1, τ_2)	0.499946	-0.000054	0.001130	(2.018, 20.073)

Notes. The table reports Monte Carlo averages over 1,000 replications. The data-generating process is the same as in Table 2, with true parameters $\omega_0 = 0.5$ and $\tau_0 = 2$. In Panels A and B, common scaling means that all payoffs are divided by 1500 at the estimation stage. In Panel C, mixed representation means that tasks 1–20 are left at their original scale, while tasks 21–40 are divided by 100. Bias is computed relative to $\omega_0 = 0.5$.