UNIVERSITY of York

Discussion Papers in Economics

No. 18/12

Rehabilitating the Random Utility Model. A comment on Apesteguia and Ballester (2018)

Anna Conte
John D. Hey

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Rehabilitating the Random Utility Model. A comment on Apesteguia and Ballester (2018)[‡]

Anna Conte[*] and John D. Hey[**]

ABSTRACT

The Random Utility Model (RUM) and the Random Preference Model (RPM) are important tools in the economist's toolbox when estimating preference functionals from experimental data. In an important recent paper in this journal, Apesteguia and Ballester (2018) cautioned decision theorists against using the RUM, suggesting that the RPM may be preferable. This short note comments on this paper, and concludes that RUM does not suffer from the drawbacks suggested in their paper.

1. *Introduction*

Following Apesteguia and Ballester (2018) (from now on A&B), we take the simplest possible story. We have some experimental pairwise choice data that we want to explain. We assume that all subjects in the experiment are Expected Utility maximisers and have a CRRA utility function $u(z)=\frac{z^{1-\omega}}{1-\omega}$ where ω is the measure of (relative) risk aversion. Now, in order to explain any experimental data, we need to have a story about the stochastic nature of the data. This is where RUM and RPM crucially differ. RPM, the Random Preference Model, assumes that the parameter ω is random, and that, once it is drawn from its distribution, utilities are evaluated without error; in contrast RUM, the Random Utility Model, assumes that ω is constant, and that a random variable with mean 0 is added to the utilities, and then the decision is taken. Let us denote the respective expected utilities by *eux* and *euy.* In RUM the random variable ε is added to their difference and the decision as to whether to choose *x* or *y* is taken on the basis of the value of $eux-euy+\varepsilon$. Take the distribution of ε to be logistic with precision parameter λ. Then we can calculate the probability that *x* is chosen for any given value of ω. The figure to the left below shows Figure 1 in their paper (with a particular value for λ), and forms the central part of their criticism of RUM.

[‡] We acknowledge very helpful comments from the authors, which have led to clarifications to this comment.
[*] Department of Statistical Sciences, Sapienza University of Rome. Email: anna.conte@uniroma1.it
[**] Department of Economics and Related Studies, University of York. Email: john.hey@york.ac.uk

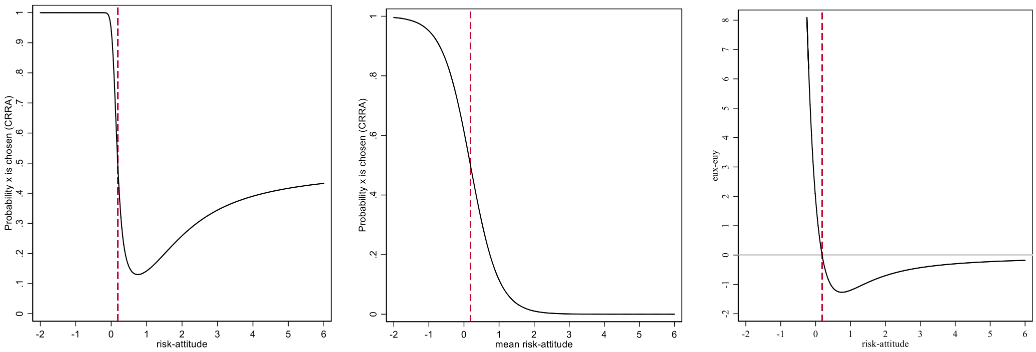


Figure 1: *RUM* *RPM* *Difference between utilities*

It will be noted that the curve is not monotonic (unlike the equivalent graph for RPM shown in the middle above), so that two values of risk-aversion may be associated with a given probability of *x* being chosen. A&B remark that this "may pose identification problems and could yield biased estimates", though we note that there is no proof of this suggestion.

Note that there is non-monotonicity in the deterministic part of RUM. Consider the graph of *eux-euy* against ω shown to the right above: it starts out positive, crosses zero, becomes negative and after reaching a trough, starts rising again and approaches zero from below[1]. This curve is non-monotonic, and implies that, after the trough, the risky option gets more and more attractive relative to the safe option as risk aversion increases, even though the latter remains favourite (the curve approaches zero from below). This is a property of CRRA preferences. We conclude that RUM captures the non-monotonicity of the preferences, while RPM does not.

Furthermore, we should distinguish between what we can logically deduce using the deterministic part and the estimates produced by maximum likelihood. In a very important sense, maximum likelihood estimation averages or reconciles the deductions from the deterministic part; this is common to both RUM and RPM.

The vertical line at $\omega=0.185$ is the value of risk-aversion at which an individual would be indifferent between *x* and *y*; we denote this by ω^*. We note that all we can logically deduce from the deterministic part is that, if the individual chose[2] *x (y)* on this problem, then his or her ω is less than (is greater than) *0.185*. The fact that the curve is not monotonic is irrelevant to this deduction, and is also irrelevant to

[1] As A&B say on page 77, "The reason is that both $U_{\omega}^{crra}(x)$ and $U_{\omega}^{crra}(y)$ approach zero with higher levels of risk aversion. Hence, $U_{\omega}^{crra}(x)-U_{\omega}^{crra}(y)$ also approaches zero."

[2] To avoid presentational clutter, we will ignore indifference throughout.

the issue of whether the model is identified. Indeed one observation is not enough to get identifiability if one has to estimate the two parameters ω and λ[3].

2. *Is non-monotonicity a real issue?*

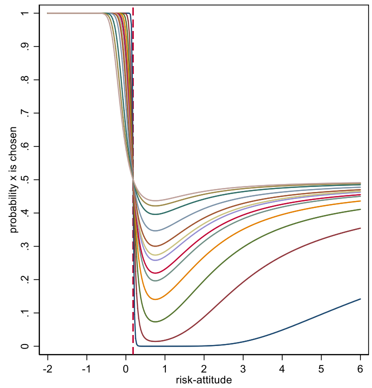


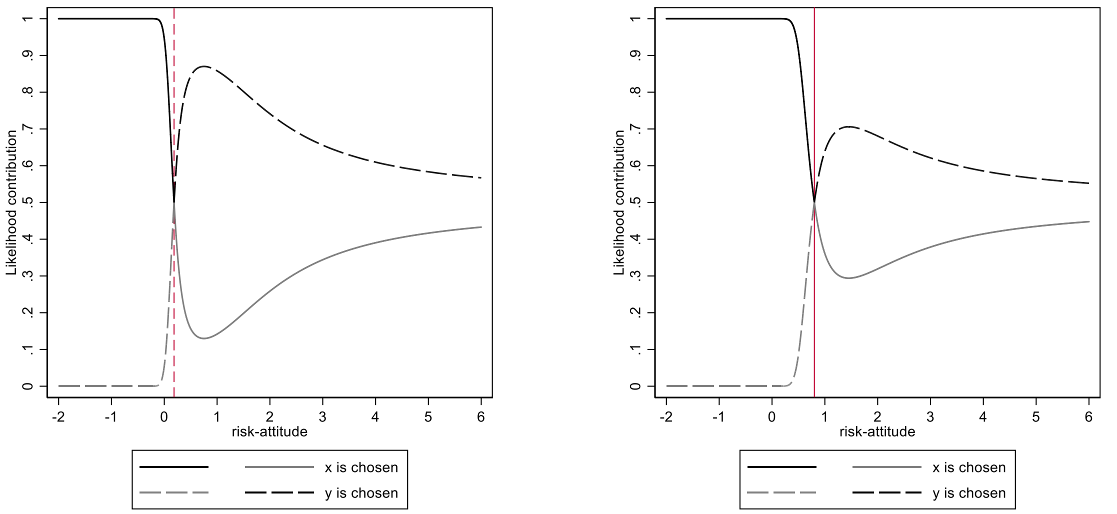
Figure 2: probabilities for several different values of λ (the precision parameter)

Let us start by considering the figure above graphing the probability of choosing the gamble *x* against the risk attitude. This corresponds to A&B's Figure 1 with lines drawn for different values of λ (the RUM precision parameter). The value of ω that equates the two, indicated by the vertical line, is the only uniquely identified point in that plot, given the infinity of equally-probable combinations of risk index and error dispersion. Therefore, a subject's risk attitude being below or above this value (depending on whether *x* or *y* has been chosen, respectively) is all we can logically deduce (from the deterministic part of either story) from this sole decision task. We note that if λ, the precision parameter, were infinite, the 'curve' would be a step function with the step at 0.185.

Having said this, let us have a look at A&B's example problems from a different point of view – from that of their contribution to the likelihood. If our target is that of estimating the individuals' risk attitude, through maximising the log-likelihood, we need to consider the information that decision tasks convey in terms of their contribution to the likelihood. The first key question is whether the likelihood function is single-peaked. If not, maximum likelihood estimates may not be uniquely identifiable. If it is, the next question is whether the true values of the parameters can be retrieved

[3] We will see in what follows that, if the value of λ is taken as given, one does get identifiability: if *x* is chosen, the maximum likelihood estimate of ω is -∞; if *y* is chosen the maximum likelihood estimate of ω is around 0.75, at the trough of the function; the corresponding RPM estimates would be -∞ and ∞. Which is more misleading? We do not know.

from the data by maximum likelihood estimation. Let us assume for the moment that the parameter λ is known[4].



decision problem A   decision problem B

Figure 3: likelihood contributions for two different problems

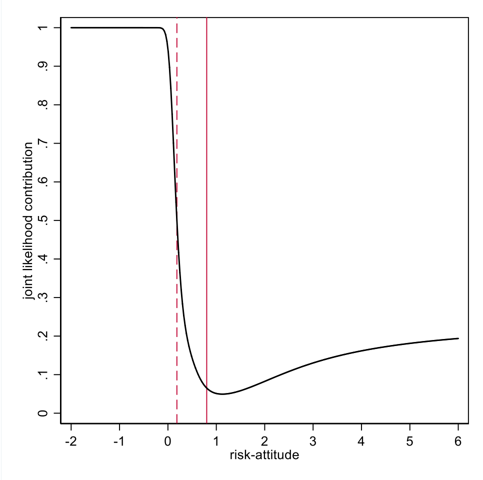
The figure to the left depicts the individual contributions that A&B's example decision problem (referred to as 'decision problem A') passes to the likelihood function to be maximised, for a given λ value, but this obviously depends upon the choice of the subject. The continuous line is the one that is passed when *x* is chosen, and the graph shows that this function is maximised when ω is at $-\infty$[5]. The dashed line is the one that is passed if *y* is chosen and this is a single-peaked likelihood function (though not necessarily peaking at the true value of ω). Ironically, it is the non-monotonicity that gives rise to this single-peaked-ness, though we are not saying that the peak is at the 'true' value of ω: all we can logically deduce from the deterministic part is that $\omega>0.185$. The figure to the right represents the contribution to the likelihood of a second decision problem identical to A&B's except for the probability of the largest outcome which is here put equal to *0.3*. This second decision problem (referred to as 'decision problem B'), has a gamble *x* less risky than in the original problem, and showing a higher ω^* at *0.801*.

Let us now look at the identifiability of risk attitude using maximum likelihood, from these two decision problems (we continue to assume for the moment that λ is known). This clearly depends on what the

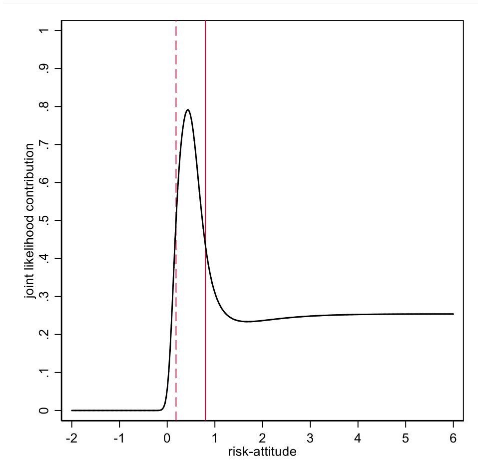
[4] We will consider later the situation where we need to estimate both (the mean of) ω and λ.

[5] As A&B have pointed out to us, even though the graph *looks* flat to the left of around -0.365, it is in fact not so and has a negative slope everywhere in this region.

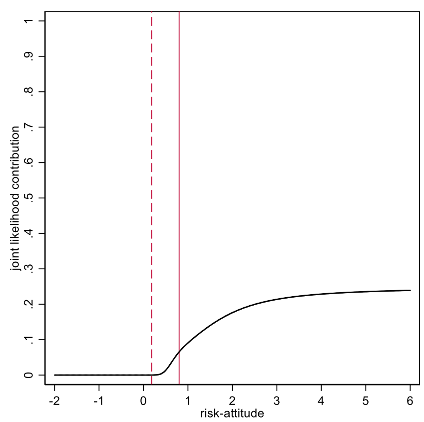
subject chooses. The two graphs below show the contributions to the likelihood depending on the choices.



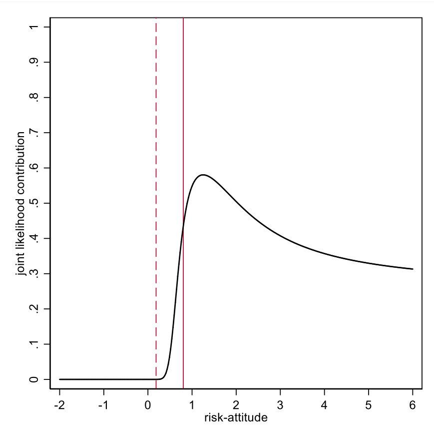
x chosen in both problems



y chosen in problem A and *x* in problem B



x chosen in problem A and *y* in problem B



y chosen in both problems

Figure 4: joint contributions to the RUM likelihood from two decision problems

There are four possibilities:

Possibility 1: *x* is chosen in both problems. The two observations lead to logical deductions from the deterministic part that $\omega < 0.185$ and that $\omega < 0.801$. Here the maximum likelihood occurs when ω is $-\infty$, which may be misleading, but is not inconsistent with what we can deduce.

Possibility 2: *y* is chosen in problem A and *x* in problem B. The two observations lead to logical deductions from the deterministic part that $\omega > 0.185$ and that $\omega < 0.801$. Here there is a well-defined unique best ω. In this case, the likelihood's maximum lies within the ω^* for the two decision problems.

Possibility 3: *x* is chosen in problem A and *y* in problem B. The two observations lead to logical deductions from the deterministic part that $\omega < 0.185$ and that $\omega > 0.801$. The two observations conflict[6] in terms of the information that they are giving, and it is not clear how the maximum likelihood estimate can be interpreted. Note that this is inevitable given the conflicting information that the two observations are giving us. There is no 'correct' way for maximum likelihood to reconcile them.

Possibility 4: *y* is chosen in both problems. The two observations lead to logical deductions from the deterministic part that $\omega > 0.185$ and $\omega > 0.801$. Here the maximum likelihood estimate is around 1.5, which may well be misleading, but if all we know is that $\omega > 0.185$ and that $\omega > 0.801$ then the estimation is not wrong[7].

In all cases, we can find estimates. This reinforces the point that it is the nature of the problems, and the decisions that the subject makes, that allow us to make inferences. We note that whether these inferences are accurate is a separate point: this depends upon whether the maximum likelihood estimates can retrieve the true values of the parameters; we will turn to this in the next section. Incidentally, we get the same kinds of graphs for the RPM specification (see Figure 6 below) and can make the same kind of inferences[8]. Note that when *y* is chosen in both problems, the RPM estimation of ω is $+\infty$, which although misleading is not necessarily wrong (we just do not know).

Now let us look deeper at this case. Recall that there are *two* parameters in the RUM model: the risk-aversion parameter ω and the precision parameter λ. The above discussion has assumed that we know the latter. However, it has to be estimated. Examine Figure 5 that graphs the RUM likelihood function (in the case of two observations in both of which *y* was chosen) as a function of ω for different values of λ (where the higher curves relate to higher values of λ). The maximum with respect to both ω and λ is clearly at $+\infty$ for both – *the same as the RPM maximum likelihood estimates* (and equally absurd from just two observations).

[6] They also conflict for RPM.

[7] In this case, the RPM maximum likelihood estimator is $+\infty$, which, while not wrong (we do not know), may be misleading.

[8] The graphs in Figures 4 and 5 are sensitive to the values of λ used. Here, we have used $\lambda = 1.5$ (as A&B) for those representing the joint likelihood contribution for the RUM model, and $\lambda = 2.5$ for the RPM model in order to make the plots easily readable in the same range of values for ω. We note that, when λ tends to $+\infty$, the likelihood contributions from the two approaches coincide.

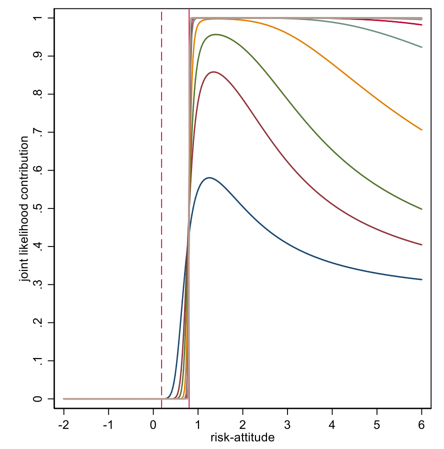
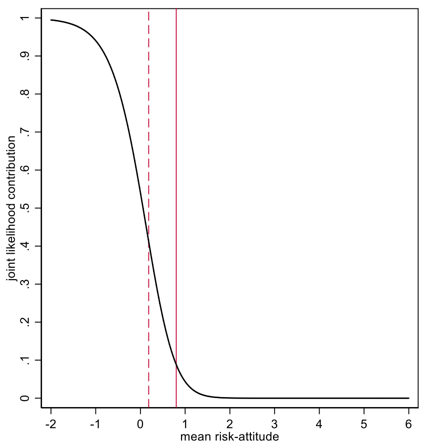


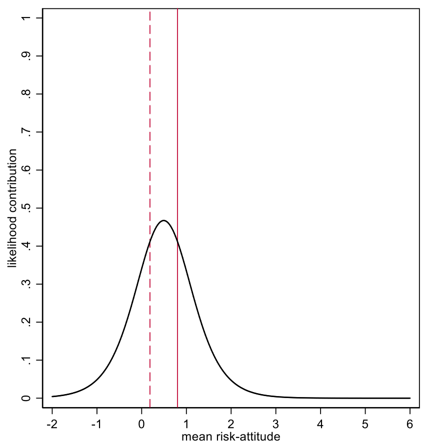
Figure 5: joint contributions to the RUM likelihood from two decision problems for the case in which y is chosen in both problems, for different values of λ

Let us now consider the problems that maximum likelihood may have with *N* observations, numbered so that the first *n* are where the individual chooses *x* and the last *N*-*n* are where the individual chooses *y*[9]. Logical deductions from the deterministic part of the stories from these observations are that ω ω_i^* for $i \leq n$ and that $\omega > \omega_i^*$ for $i > n$. If it happens to be the case that the ω_i^* for *i>n* are all smaller than the ω_i^* for $i \leq n$, then there is no problem, no conflict in the information given to us, and we can safely conclude that ω lies between $\max(\omega_i^*; i > n)$ and $\min(\omega_i^*; i \leq n)$. Moreover, the maximum likelihood estimates for both RUM and RPM will be in this interval. However, if there is *any* value of ω_i^* in the set *i>n* larger than *any* value of ω_i^* in the set $i \leq n$, then there is a conflict in the information provided. Maximum likelihood estimation must resolve this conflict somehow, both for RUM and RPM. How this is done depends upon the shape of the various likelihood functions, but non-monotonicity itself does not change the nature of the problem, though it clearly affects the solution. It *may* imply biased estimates, but this is an empirical issue. We investigate this in the next section.

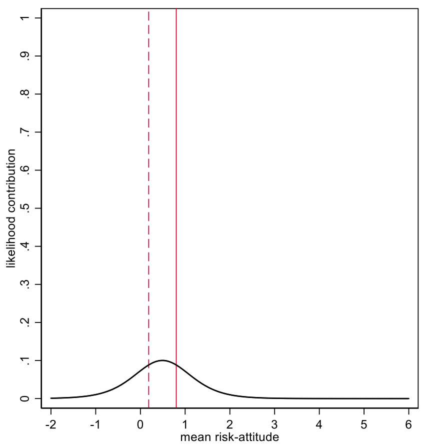
[9] We shall ignore the uninteresting cases where the individual either chose *x* everywhere or chose *y* everywhere.



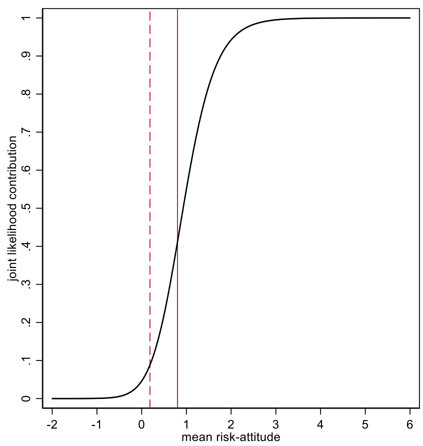
x chosen in both problems



y chosen in problem A and *x* in problem B



x chosen in problem A and *y* in problem B



y chosen in both problems

Figure 6: joint contributions to the RPM likelihood from two decision problems

3. *Identifiability and retrievability*

We use the data from Andersen *et al* (2008), which is the same data that A&B use. We start by presenting the two figures below. The one to the left shows the RUM log-likelihoods for the data set plotted against ω for a number of different values of λ; the one to the right shows the log-likelihoods for the data set plotted against the value of λ (the precision parameter in RUM) for a number of different values of the risk attitude ω. It is abundantly clear that all the curves are convex and have a unique maximum. So the model is identified.

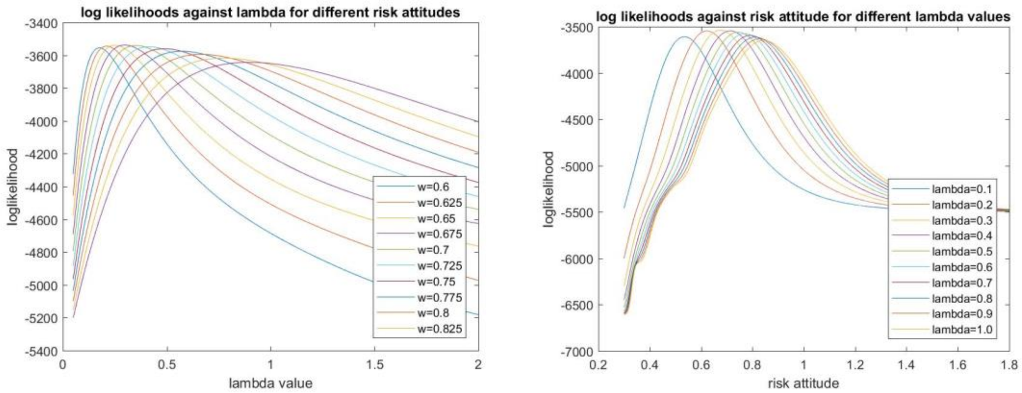


Figure 7: RUM log-likelihoods plotted against ω and λ

In a sense, this is implicit in the fact that A&B report the results of estimations using the same data for both one version of RUM and one version of RPM. They could not have done this if the models were not identifiable. But this rather begs the point as to whether the estimates retrieve the true values. To explore this latter point, we take the same set of problems, but, instead of estimating on the actual data, we generate 1000 repetitions, taking their estimated parameters to be the true parameters, estimate[10] for each repetition and take the means and standard deviations. The results are below (with the standard deviations below their respective estimates).

true		estimated	
		RUM	RPM
RUM	.661	**.661** .0116	.696 .0151
RPM	.752	.676 .0151	**.752** .0157
mean (mean) parameter values			

true			
		RUM	RPM
RUM	.275	**.276** .0264	3.267 .1673
RPM	2.495	.343 .0496	**2.498** .1031
mean dispersion parameters			

true	Estimated	
	RUM	RPM
RUM	**-3491.8**	-3571.3
RPM	-3562.8	**-3467.5**
maximised Log-likelihoods		

It will be seen that when the estimated model is the true model, then the mean estimates are close to their true values, but when the estimated model is *not* the true model, this is *not* the case. Also the (maximised) log-likelihoods are the largest when the estimated model is the true model. This seems to indicate that both RUM and RPM are identified and that the parameter estimates for both models retrieve their true values from the data.

In passing we should note that, just because A&B's estimated parameters for the RUM are lower than those for RPM, this does not imply that the former are biased downwards. We cannot know as we do not know the truth. In any case the fixed value of the risk-aversion index for RUM is a different thing

[10] Incidentally, as the eagle-eyed reader will have noticed, there is also estimated a tremble parameter, as the RPM cannot cope with observations for which the subject chose a dominating option. For the sake of fairness we introduced a tremble for both models. The trembles were estimated precisely in all cases.

from the mean of the distribution of the risk-aversion index for RPM. The stories being estimated are simply different.

4. Concluding comments

Before concluding, we should note that, while A&B recommend the use of the RPM, its application is not without its difficulties. For the case we and they have considered, where there is just one random parameter, estimation of RPM is straightforward: for each problem we just need to calculate the value ω^* at which the subject is indifferent between x and y and then calculate the probability of ω being greater or less than that. If, however, there are, say $n>1$ random parameters, one has to find the frontier (in n-dimensional space) over which the subject is indifferent, and then calculate the probability of the (n-dimensional) parameters being inside that. Based on the existing literature and on our own experience, this is not always an easy task. In contrast, RUM is straightforward – though we agree that this is not a reason for its use.

Be that as it may, we have shown that both models are identified and produce estimates of the true parameters close to their true values. It follows that we have shown that it is safe to use the RUM when estimating preference functionals from experimental data. The non-monotonicity of the probability function does not create problems.

References

Andersen S, Harrison GW, Lau MI, and Rutström. EE (2008), “Eliciting Risk and Time Preferences.” *Econometrica* 76, 583–618.

Apesteguia J and Ballester MA (2018), “Monotone Stochastic Choice Models: The Case of Risk and Time Preferences”, *Journal of Political Economy*, 126, 74-106.