Review of: "How can we Guess the Distribution of Wealth in Greenville, Mississippi? A Proposed Heuristic for Guessing Distributions"

Robert Veszteg[1]

1 Waseda University

Potential competing interests: No potential competing interests to declare.

While I find the proposed research question interesting, I do not believe that this article is publishable in any scientific outlet. It is not only that its style and format are unusual, but it ultimately fails to deliver a convincing result.

1. Format

- There are just too many words, expressions, and even sentences directly quoted from somewhere else, sometimes even without a proper reference (e.g., the sentence right above Equation 3).
- Equations and figures need to be formatted so that they have the same style (and font size, at least).
- Note that the parentheses in Equation 6 do not match.
- Also, it is unclear to me why some words (like median, binomial, multinomial, guesstimating) are capitalized.

2. Mathematical notation

- The three key pieces of notation *N*, *n*, and *r* are not used consistently throughout the text. This is serious mistake that confuses the reader. As they first appear, they seem to carry the following meaning: *n* is the number of quantifiers or categories (p.2), *r* is the number of instances in the decision-maker's sample (p.2), *N* is the number of possible combinations given *n* and *r* (line 4 on p.4). Later, however, *n* appears as the number of *N* measurements that are larger than the true median (line 9 on p.4).

3. Content

- The authors claim that the "proposed heuristic" for guessing distributions is *straightforward* (p.1), *clear* (p.1), and that it works *so well* (p4. and p.9), but they fail to clarify how they measure its performance. Similarly, they talk about *successful guesstimation* (p.4), *choosing the best distribution* (p.4) and *less biased distribution* (p.4). Nevertheless, they never explain what is to be considered as *successful*, *best* and *biased* in their framework.
- There seem to be (at least) two important assumptions behind the claimed results that remain in the background. Their role must be acknowledged, emphasized and carefully described, especially because the authors claim to deliver a method that is *the best*.
    - Assumption 1: the sample that the decision-maker works with has independent measurements (p.4). This is crucial

for Equation 3 to be true. I wonder how this assumption is defendable when the decision-maker is allowed to be biased and boundedly rational in the model.

- Assumption 2: the proposed method relies on a hypothesized distribution (p.4). Again, this is crucial for the (true) median to be estimated correctly. How is it possible that the decision-maker is familiar with probability distributions (e.g., Pareto), while is unable to think in professional terms when it comes to architecture and income inequality.

- An additional concern of mine related to the claimed performance of the proposed method is that the numerical tests included in the text (p.5) do not warrant the general result stated by the authors. In other words, it remains unclear how the method works in general and how robust it is.