MS. NO. PSRM-OA-2019-0126 ENTITLED "TAKING DYADS SERIOUSLY"

Dear Editors,

Thank you for the opportunity to revise and resubmit our manuscript. We believe the manuscript has benefited from the Reviewers' helpful and thoughtful comments. We have revised the manuscript, taking seriously each individual point raised by the Reviewers. Our comments and responses are shown in *BLUE* below each of the points made by the Reviewers.

We hope you agree that the manuscript has improved through this process and we are looking forward to your response.

Reviewer 1

(1) P1: "Most standard approaches assume that the problems raised by non-iid relational data can be addressed by recalculating the standard errors of estimated parameters to reflect the potential clustering of cases. In practice, such strategies rarely work because they do not directly address the fundamental data generating process".

- The term "rarely work" is too casual. What problem exactly is motivating this paper? E.g., the Aronow et al. paper cited at the start of the same paragraph looks at consistent variance estimators for semiparametric regression estimators on dyadic data. By the standards set out by those authors in that paper, the strategies "work" just fine. I suppose then that the current paper has a different sense of what it means for something "to work." For example, I am supposing that the current paper is talking about people misconstruing the structural interpretation of a reduced form parameter estimated in a GLM fit to dyadic data. That would be wholly orthogonal to what Aronow et al. were considering, but then this needs explication.

- *TO DO in paper: Strategies like those employed by Aronow et al do not directly model all relevant dependencies in the data* [1] *but instead adjust the variance estimator. While such an approach decreases the risk of a type 1 error, it increases risk of type 2 error (if the mean estimate is biased downwards, increasing the variance estimate will cause us to falsely accept the null). The main aim of our research is to directly account for the relevant dependency structures in the data, thus providing an accurate estimate and addressing both types of possible errors.*

(2) P2: "Scholars working with dyadic data typically begin by stacking observations associated with each dyad on top of one another. This makes sense if each observation is independent of the others."

- Revise this opening, since it is perfectly fine to stack observations and then incorporate auxiliary information (such as adjacency matrix) to characterize dependencies between rows.

- *The point is well taken, we have clarified this section (on page 2) to point to the assumption of independence as the primary problem, not the stacking of observations in and of itself.*

(3) P3: "Unless one is able to develop a model that can account for the variety of explanations that may play a role in determining why a particular actor is more active than others, parameter estimates from standard statistical models will be biased."

- This is another example of language that is too causal and imprecise, like the opening paragraph. What are "parameter estimates from standard statistical models"? And "biased" with respect to what target of inference? Again, I think the authors have in mind "bias" with respect to parameters that are meant to have a particular structural interpretation, but this needs to be clarified.

- Generally, section 2 should be clearer in motivating the problem. Provide us with a toy example that illustrates the kind of "bias" that the paper has in

---

[1]As Aronow et al explain, "Of course, accounting for dynamic dyadic clustering may fail to fully account for all relevant dependencies in the data." p. 16.

mind from "standard models". For me at least, after reading section 2 I still wasn't sure what problem the paper is actually trying to address.

- *MG/CD: We think the reviewer makes a good suggestion here to clarify, mathematically or in some other illustrative way, exactly how we think that dependencies bias the estimator. For example we can say that our target of inference is the estimation of beta (with some variance). If a coefficient is correlated with a dependency, then we could obtain an estimate for beta that is artificially high (and the same is true for the reverse).*
  - *What we think: True DGP: $Y = B0 + XB1 + a + b + $ 3rd order stuff $+ e$*
  - *What most do: $Y = \hat{B}0 + X\hat{B}1 + e$*
  - *Something like if we have dependencies then $E(\hat{B}1) = E(B1) + cov(B1,$ other stuff)*

(4) P7: Regarding the specification, this augments a dyadic random effects model, which is quite common, with the LFM component, which resembles an interactive fixed effects specification (cf. Bai 2009, Ecta) and is based on an SVD factorization of the observed network.

- It is useful to think of the infeasible estimator and the implications for interpreting the LFM component. The infeasible estimator takes the transformation f(.), fixed effects parameters, and SRRM random effects as know, in which case the LFM amounts to predictions from a factor model fit to the adjacency matrix of residuals on the scale of the latent outcome. Would this be a reasonable interpretation? If so it would be useful to draw the connection to factor models commonly in use these days (such as the interactive fixed effects model of Bai, which is the basis of, e.g., generalized synthetic control methods).
- *Insert great response.*

(5) P8: "accounting for this structure is necessary if we are to adequately represent the data generating process." Continuing with the problems I have with the language above, what does "adequately represent" mean?. P9: "a) biased estimates of the effect of independent variables, b) uncalibrated confidence intervals, and c) poor predictive performance."

- Okay here is where the paper are starting to clarify goals. So a) is about causal identification. But as I understand, the solution here is parametric identification of causal effects. This means that the issue of misspecification and robustness ought to be taken up. Similarly, b) is addressed in this paper under a fairly simple parametric data generating process. But again, how robust is the proposed model to situations where the real DGP does not abide by the parametric assumptions (linearity, normally distributed random effects, and a factor model for the residuals of dimension K)? For c), I think most would agree that the "proof" is simply in comparison to whatever other good models are available. Not however, that in social science applications, forecasting and therefor c) are rarely of interest (at least not in the types of research that appear in political science journals). As such, from the perspective of what the researchers in discipline mostly cares about, a) and b) are of primary importance.
- *Insert great response.*

(6) P9-10: we need to know exactly how and from what distributions X and W are being drawn to interpret the results here. Are they independent/orthogonal or no?

- *SM will deal with this.*

(7) Section 4 needs to incorporate an assessment of robustness to misspecification. See the Aronow et al. paper for an example of how to do this, where they study the performance of a dyadic random effects model.

- *SM will deal with this.*

Reviewer 2

(1) The additive and multiplicative effects (AME) model presented in the paper allows for a better treatment of first, second and third order dependencies in the data, and account for unobserved sources of bias in the data. The presentation of the AME estimator is clear, and the simulations provide clear evidence that the model outperforms naive OLS/GLM models fitted to the same data. The paper will be of great interest to the PSRM audience and is likely to have a strong impact on future research in international relations. I do, however, have a few suggestions for revisions that will hopefully strengthen the paper's contributions.

- *We thank the reviewer for their comments and suggestions.*

(2) The paper shows that the AME model performs well in comparison to the "oracle" model where all sources of dependencies in the data are accounted for, which is reassuring. Yet this comparison feels incomplete. First, the naive model is clearly misspecified given the underlying data generating process. A large number of dependencies in international relations are linked to geography. What would happen if the dependencies were modeled using spatial econometric techniques?

- *Insert great response.*

(3) It would help if the author discussed in more depth the properties of the AME estimator in the presence of unobserved dependencies. One would imagine that the oracle model, where all the unobserved factors and dependencies are accounted for should always outperform a parametric or semi-parametric estimator. Relatedly, it also would help to have a more technical discussion of the properties of AME models, particularly in terms of the bias-variance trade off.

- *Insert great response.*

(4) The author could also underscore a related issue, which usually escapes empirical IR scholars: the importance of developing theories that account for the direct and indirect effects of interactions in a setting where the units are interconnected, and where the effects on one unit have the potential to impact other units in the system. A traditional approach is to throw as many observable variables as possible as controls, without much consideration on where they fit in the model. As the comparison of the AME and oracle models suggests, a careful modeling of spillovers and dependencies can result in better estimates of the underlying relationships.

- *TODO/CD: We agree with this suggestion. We have now emphasized the importance of developing theory in our discussion of the AME and Oracle comparison (p. X), as well as in the conclusion of our paper (p.Y).*

(5) The paper emphasizes the importance of estimating the random effects for theory building and hypothesis test. A longer elaboration and illustration of this point would be very helpful.

- *TODO/CD: I think this comment and the above comment are essentially the same. We could even cite our JOP as an example of incorporating theory? Perhaps other new-ish net paper as well?*

(6) I would like to see a more detailed discussion of the setup and choices made for the Monte Carlo simulations. What type of network do the simulated data model? How does the network in the simulation relate to typical networks in IR studies, and to the ones in the replicated papers.

- *SM will deal with this.*

(7) The criteria for choosing the studies for replication is unclear and needs a better justification. I would have liked to see a larger set of replications particularly in areas and topics where the data generating process, the dependencies among units varies, or the relationship between the units is hierarchical, cross-classified or multilevel.

- *TO DO/CD: We selected our cases based on a few criteria. We focus on studies that are explicitly about International Relations, were published in the last 10 years, and were published in a top ranking general political science outlet (for consistency in editorial standards and reviews, we also decided to focus on one journal, the APSR). We hope that these three criteria ensure that our paper is readable and interpretable to an applied audience. In addition, we selected studies that could help illustrate different key components of the AME approach (Max, more here on each?). While we agree with the reviewer that further replications would effectively demonstrate the benefits of the AME, they are beyond the scope of this paper at this time. (CD: Should we also mention something like 'in consultation with the editor, we believe they are beyond the scope....")*

(8) Given the importance of promoting the use of the AME method I would encourage the author to revise the tutorial presented in Appendix C to make it more accessible to a larger audience.

- *The reviewer raises a valid point. We have expanded our tutorial in Appendix C, particularly in giving more detail and concrete examples when it comes to formatting the data, which may represent the biggest hurdle for some users. We also...*