

Dear Prof. Minhas:

Manuscript ID PSRM-OA-2019-0126.R1 entitled "Taking Dyads Seriously" which you submitted to Political Science Research and Methods, has been reviewed and I have read it carefully as well. The comments of the reviewer(s) are included at the bottom of this letter.

I invite you to respond to the reviewer(s)' comments and revise your manuscript. I will not ask the reviewers to do another round, but I would like you to take their (particularly reviewer 1's) concerns seriously. R1 continues to have problems with the language of the manuscript. I would like you to try to be as precise as possible as to the manuscript's contribution. For example, in the first sentence

"The aim of this work is to address how to estimate the effect of exogenous covariates in a GLM context when there are network dependencies between observations in dyadic data."

"effect" is not precisely defined, so you might substitute "how to estimate regression coefficients in a GLM context."

Also, my understanding is that modeling the "underlying structural features in the data" means including these "structural features" in the conditional mean equation of the GLM rather than including them exclusively in the variance equation (or ignoring them altogether). Or, more precisely, including SRRM and LFM parameters in the conditional mean equation. The advances you make come from this specific modeling choice. Augmenting the GLM in this particular way leads to 1) unbiased (or nearly unbiased) estimates of regression coefficients under homophilic dependence, when these coefficients exist and are known (the Monte Carlo simulations) and 2) superior out-of-sample forecast predictions for real data, in which case the regression model is just an approximation with unknown parameters (the replication exercises). Including LFM parameters in the conditional mean equation also makes it more likely that IR scholars will treat dyadic dependence as substantively important.

These are just suggestions for ways to make the language more precise. R1 has some specific suggestions too, as well as two or three additional requests about the model and simulations. R2 has a couple of suggestions for the replications. After you have revised your paper, please write me a short memo that describes how and where you have addressed these concerns.

To revise your manuscript, log into

https://urldefense.com/v3/_https://mc.manuscriptcentral.com/psrm_!!HXCxUKc!jG7d_HbaWdLbAmd8ob1E4Ti0fZW

[2LkNYNKNbwdfk6a0Fn1H93nZyM_FlhMOQ794\\$](#) and enter your Author Center, where you will find your manuscript under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number will be appended to denote a revision. You may also click this link to start your revision:

*** PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. ***

[https://urldefense.com/v3/https://mc.manuscriptcentral.com/psrm?URL_MASK=ae1df591022246a1960b5f817b1008ac;!!HXCxUKc!jG7d_HbaWdLbAmd8ob1E4TlOfZW2LkNYNKNbwdfk6a0Fn1H93nZyM_FlOY8lnwo\\$](https://urldefense.com/v3/https://mc.manuscriptcentral.com/psrm?URL_MASK=ae1df591022246a1960b5f817b1008ac;!!HXCxUKc!jG7d_HbaWdLbAmd8ob1E4TlOfZW2LkNYNKNbwdfk6a0Fn1H93nZyM_FlOY8lnwo$)

When submitting your revised manuscript, you will be able to respond to the comments made by the reviewer(s) in the space provided. Please use this space to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).

Because we are trying to facilitate timely publication of manuscripts submitted to the Political Science Research and Methods, your revised manuscript should be uploaded as soon as possible. We expect to receive your revision by 14-Feb-2021. If it is not possible for you to submit your revision by this date, please contact the Editorial Office to rearrange the due date. Otherwise we may have to consider your paper as a new submission.

Once again, thank you for submitting your manuscript to the Political Science Research and Methods and I look forward to receiving your revision.

Sincerely,
Prof. Jude Hays
Associate Editor, Political Science Research and Methods
jch61@pitt.edu

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author

Thank you for the chance to evaluate the revised manuscript. While I appreciate the revisions already made, I think more work is needed for this paper to have real impact. This includes, ironically, backing off of some of the more general claims, and then working through more details.

1. Opening section

In my previous review I pointed out some problems with vague language, mostly in the opening sections. Honestly, I still find the introduction to be problematic. Maybe it could be cut and we just open with the part of section 2 where the paper begins to discuss processes that lead to dyadic and higher order dependencies? The reason is that the intro tries to strike a very general tone. But in reality, the paper proposes a specific estimation strategy for parameters in a specific parametric model of networked interactions. I think the paper would be more useful to readers if it started with the specific applications, and then on the basis of that, motivated the statistical specification on substantive grounds. There are some swipes at other literature are neither necessary nor clear enough to be very compelling. E.g., consider these parts of the intro:

"Most of these studies use a generalized linear model (GLM). However, this approach to studying dyadic data increases the chance of faulty inferences by assuming data are conditionally independent and identically distributed (iid)."

It should really read, "Most of these studies use a generalized linear model (GLM). However, this approach to studying dyadic data increases the chance of faulty inferences **if the estimation and inference assume** data are conditionally

independent and identically distributed (iid)." But then if this is all just about mistakenly assuming iid, just say it quickly and move on.

"In practice, such strategies remain insufficient because they do not directly model underlying structural features in the data but instead adjust the variance estimator." with footnote "While this 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)."

Why would the mean estimate be biased downward? And, ****biased with respect to what target of inference****? A problem with the opening discussion is that it does not establish a target of inference. Without doing so, "bias" is undefined. It seems like they mean the coefficient for the linear relationship between some regressor and the latent outcome variable. But what if one is using a nonparametric model or a semiparametric model that operates on a different latent space---is bias even defined? It seems that for bias to be defined we have to be restricting ourselves to a class of models that propose the same latent space, but that makes this much less general that the discussion leads on. Moreover, that paper is not committing to talking about bias relative to causal effects under effect heterogeneity, but isn't that usually what people are after? Or, maybe the paper is really concerned with bias with respect to a forecast prediction? If so, then we need to set up a forecast prediction error analysis.

I actually don't think it's really worth it for the paper to try to sort these issues out. Rather, just move on to the substantive problem, and pitch your (semi-)parametric representation of it. Tell us how it is better than what others have done, and don't worry about trying to make this more general than it really is.

Moving to section 2, the extended discussion of specifying a likelihood under iid is overkill. Again, I would just move on to the discussion of the application. The point that iid is inappropriate will receive due attention in the course of developing the networked interaction model.

2. Explanation of the model

The references to the recommender systems literature and the work of Bai and Pang are appreciated.

Explain how inference is done for the parameters from the AME model. How are intervals produced exactly? (Need this because you examine these in the simulation and make many statements about intervals throughout the text.) What are the operating characteristics, in terms of convergence properties, etc.?

3. Simulation

The DGP is simple enough that the paper needs to include the analytical calculation of the bias from the simple probit regression of Y on X. We get no insight by just having monte carlo simulation results.

For the simulations, clarify the "standard international relations approach" as just Y regressed on X, when you introduce it. Then explain how confidence intervals were produced. Are these just normal-approximation Wald intervals with non-robust standard errors?

Explain exactly how the confidence intervals are produced for the AME model. They were never characterized in the section that discussion the model.

Reviewer: 2

Comments to the Author

This is an excellent paper. It provides an alternative solution to GLM to address problems found in modeling dyadic relations in international relations and other fields where the data generating process includes first, second and higher order dependencies. The paper is well written and the AME modeling process is clearly presented. The tutorial in

particular is easy to follow, which would result in more scholars adopting the strategy.

I do have a couple of minor comments, some of which would be easily addressed, or ignored:

1) I would have liked to see a justification for the choice of papers used in the replications. It is still unclear. Are these cases where the theory fail to characterize the direct and indirect effects, and implicitly assume that the latent dependencies are nuisances?

2) The presentation of the AME model clearly states that the X s are exogenous; it is likely that they are not in replications. Still, the AME estimates (which in the simulations do not seem to outperform the standard method for estimating the betas when correlation between X and W is high) suggest that the models in the original papers are misspecified, and that these differences are likely driven by a better modeling of the DGP. This is a very important take-home point.

In any event, I believe this is a great paper. I look forward to see it in print.