This is a difficult manuscript to review because the author has some good ideas, but sells them in a way which is at best misleading and presents other arguments which are partially correct and partial incorrect and still others which are simply exaggerations used to sell the paper. The reviewer has to resist the temptation to write a long and involved review separating out the good from the bad, correcting the correctable and noting the hopelessness erroneous statements. But that would be doing the author's work for her.

Detailed comments:

"Under our inferential framework, analysts merely add a simple preprocessing step to their usual data analysis procedures. They then employ whatever statistical models they are accustomed to using to analyze the preprocessed data rather than the raw data. All of the intuition, diagnostics, and knowledge about our parametric procedures can then be used as before."

These statements are from the introduction but similar statements appear throughout the paper. This is one of the main claims of the paper and it is incorrect and horribly misleading. As the authors know very well, current practice in the social science involves estimating parametric models with k covariates where k is much larger than 1 and interpreting each and every estimate of these k covariates as causal. In essence social scientists live in the world of structural estimation. For example, researchers model and not leave out post-treatment variables.

The authors of this paper suggest something which is a *radical* departure from standard practice and they appear to be afraid to admit this. Are they afraid of scaring away potential customers? The authors want consumers, I mean researchers, to preprocess their data with matching where one of the k covariates becomes the treatment variable of interest. I assume they want the other k-1 usual covariates to become controls instead of treatment variables. On this point, I think the authors suggestion makes sense, although it is not innovative in any way, but it is radical.

Further issues are also radical. For example, the concern about post-treatment bias (p.31). They have to change Carpenter's model before they are willing to estimate it! How's that for just doing what you always do and not having to change standard practice aside from a little preprocessing.

"To our knowledge, the present paper is the first to propose and work

out the conditions for matching as a general method of nonparametric preprocessing, suitable for improving any parametric method."

Aside from one controversial issue noted below, I have no idea what is so new about what the authors are proposing other than the sales job involved. Others have used a wide variety of parametric and semi-parametric models. The authors should review the literature on the subject.

"Connections to Missing Data and Ecological Inference"

I agree that causal inference is very difficult and involves unobservables, but this analogy to EI doesn't make sense to me. I don't know what it is adding and the claim that causal inference is a particularly difficult case of EI is not supported. May be it is a correct claim (although I doubt it), but they don't examine the issue enough to have the claim in the paper. Some causal effects, say in benchtop science, are easy enough to estimate. And so are some EI where the bounds are highly informative. The issues rests with the other cases which require analysis.

"Random Causal Effects"

This is the relatively new things the authors bring to the table in the discussion of parametric bias adjustment. This isn't a new idea as in no one has said it before, but it is new in the sense that it is a sharp departure from what is usually assumed in the literature. The idea of a random causal effect introduces another disturbance (aside from the one induced by the random process of treatment assignment). This extra random component is a sharp departure from the Fisher setup as well as the Neyman setup coming down to Rubin through Cox and Cochran. Whether this idea which brings in a lot of the metaphysics of structural estimation is worth it is not discussed by the authors but it should be.

This is a minor point, but the authors talk about conditional vs unconditional estimands but is not clear to me what relationship these have, for example, to population and sample estimands. Are population estimands possible? Are they also random (aside from treatment assignment)?

"Finally, we note two advantages of our general framework. First, our two-step procedure is doubly robust in the sense that if either the

matching or the parametric model is right, causal estimates will be statistically consistent (see Robins and Rotnitzky, 2001). That is, if the parametric model is misspecified, but the matching is correct, or if the matching is inadequate or wrong but the parametric model is correctly specified, then our estimates will still be consistent. The double robustness property does not apply generally to using matching with a difference in means."

This is just one of the claims the authors make which is wrong and I imagine included only to sound current. Counter examples are trivial. For example, the authors' preprossing method involves dropping observations without common support (p21). Imagine that the parametric model is correct. In an extreme example matching could leave the analyst without any valid observations (because of support problems) and hence no estimate from the (correct) parametric model. There are less extreme examples. Matching could drop the observations which are required to identify a non-linear functional form or simply drop observations which would make the non-linear functional form change shape etc.

There are cases where post matching adjustment bias adjustment can lead to increase variance and bias (where, for example, there are inliers). The authors should talk about this and how researchers should decide about the tradeoffs. There is a literature here. They should engage it.

"The Balance Test Fallacy"

There is a lot of useful material in the section but the authors fail to note that we only match in order to reduce bias and bias is a function of, for example, the mean difference and variance thus looking only at the mean is not sufficient even in the linear case. See the equations for hidden bias to see what is at play. You may be better off with bigger mean differences (in terms of bias) depending what is going on with the other moments.

"Our idea is to take advantage of a common feature of all of the methods of computing uncertainty estimates associated with parametric methods: They are all conditional on the pretreatment variables X (and t), which are therefore treated as fixed and exogenous."

The approach of treating X as fixed may have some appeal, but it is not true that all parametric methods treat X as fixed. Note the parametric versus non-parametric bootstrap. One samples over the

distribution of X and the other does not.

Since the authors have assumed two disturbances (random Y and random T), their setup starts to look a lot like trying to estimate 2SLS without taking into account the uncertainty of the first stage. Supporters of matching always tell me this is okay, but I just don't see why this is okay. Tell me how this is different from 2SLS?

Finally, and very importantly it would be very helpful if the authors provided a way of measuring model dependence since this is the ultimate focus of the paper. Rigorous sensitivity tests for model dependence a la Rosenbaum and Rubin would be useful.

And equally important, if the authors are selling this method by saying that it reduces modelling assumptions it would be useful to compare their method with the usual semi-parametric alternatives that social scientists know well.