

Journal: Political Analysis
MS: PA 70-188
Title: Matching as Nonparametric Preprocessing...

This paper has a lot of interesting things in it, and a revision should be published, but it needs to be tightened up and beefed up before publication. The general topic of causal inference and the specific focus on matching is very interesting and current right now, and so the *PA* readership will be interested. The authors are also marketing some software that looks interesting, and so that is another attractive aspect of the paper. But the paper tries to do too much, isn't as focused as it should be, and is not as rigorous, clear or compelling as it should be on some important points.

The authors say they want to reduce the way that causal inferences in parametric analysis are "model dependent". Their proposed solution is to use matching to restore/transform the data to a state where it (approximately) looks the same as data we'd see from an experiment, in which random assignment to treatment and control groups was successful: i.e., receipt of treatment ($t_i = 1$ if subject i is treated, $t_i = 0$ otherwise) is independent of (and hence uncorrelated with) covariates X_i , and the assessment of the treatment effect can be as simple as computing a difference of means $\bar{Y}_i|(D_i = 1) - \bar{Y}_i|(D_i = 0)$, or (as the authors propose) via a regression of Y on t or some other kind of "standard analysis", with X included as a set of controls. In this way the methods proposed in the paper fall into the large and growing class of matching-based methods for causal inference. So far so good.

One of my big gripes with the paper is to do with the concept of model dependence as a driving, organizing theme for the paper. We all know about properties of estimators: bias and sampling variance, consistency and so on, all of which we talk about conditional on a specific model. And of course, I change the (parametric) model by changing the set of control variables X , I get different estimates of the effect of t_i , all of which may be conditionally unbiased and even efficient *given the particular model*. These estimates vary as model specifications vary. There are tools for dealing with this that are well known in the literature, as the authors indicate in a footnote: e.g., Bayesian model averaging and Leamer's EBA. One of the nice things about Bayesian model averaging is that the uncertainty associated with the parameter estimates given each model is also taken into account when forming the final, model-averaged summary of uncertainty over a parameter: it would be a nice final setp in the author's proposal if they could also do this (i.e., sampling uncertainty in each parametric model based estimate of a treatment effect contributes to the final, model-averaged estimate of a treatment effect and uncertainty in that estimate; perhaps the authors already do this, I'm not sure, it wasn't clear from the paper).

The authors need to do more work on a *really important* issue that they currently seem to skirt around: the bias and variance tradeoff underlying all statistical modeling, matching in particular, and especially for their proposed estimator. It is an ongoing intellectual scandal that there must be know 18 different proposals for estimating causal effects after matching

(of various guises), but precious little theory re the properties of these estimators and even a controversy as to how to compute standard errors. The fact that this is still an unsettled field permeates the current paper (see specially Section 6): there are lots of references to “other approaches” in the statistical literature, lots of remarks that have a “you could do this, or you could do that” flavor, and not enough comparisons helping us understand where what the authors propose falls into the panoply of approaches out there, or the tradeoffs involved. It might be genuinely helpful to provide a table or some kind of summary of other approaches, and even explicit comparisons for the examples, or Monte Carlo evidence etc.

I have to stress that this is really important for the present paper, since the authors are proposing tossing away data where we don’t have good matches, and hence they have to take a hit on variance. To say, “ah, but we have less variation over all possible subsets of regressions” dodges the issue, since I either have to choose one (or more) of those regression estimates as my estimate, or engage in some kind of averaging over those specifications, in which the sampling variance of each is implicated. So, to the authors I guess I’d say the following: “come on, you’ve either got an estimator of ATE/ATT that has low MSE relative to competitors or you don’t, lets see what you’ve got...”

As the authors think about revisions, I would also stress that the *PA* readership and political science more generally could benefit from a gentle yet rigorous walk through the issues at stake here, with explicit ties to theorems and propositions etc from the statistics literature. Section 6 is a little too breezy, or should come earlier? Part of the problem here could well be organization: with “model dependence” as the stalking horse, properties of estimators/matching procedures comes much later in the paper, and in a way that is just too loose to be helpful to someone who isn’t already familiar with a lot of the language, definitions and concepts. Re-stating some key results (theorems/propositions stated without proofs?) might be a step in the right direction? All this is geared at getting the discussion oriented and helping users understand what is stake in the choices among rival matching-based estimators (or estimators that don’t use matching), and making those choices intelligently, say on good old MSE or even asymptotic MSE criteria...?

Let me also say that the motivating example in Figure 1 is unconvincing. No data analyst worth their salt would ever fit y as a quadratic function of x for both T and C . Quadratic looks like it works for C , but linear within T . Further, why parametric fitting only? Indeed, elsewhere at least one of the authors has championed neural nets, and “local (parametric) fitting” of some kind may well dampen the effect of outliers: in fact, what the authors are proposing is actually a “human-guided” form of local (linear?) fitting, with cases outside a prescribed neighborhood getting zero weight.

Finally, I was looking for, but didn’t see, a clear, step-by-step statement of the authors proposal. This isn’t good enough. Could we have this in a table, or as a “pseudo-code” algorithm?

More specific gripes:

- What, precisely, is the problem with *model dependence*? Can this be formalized?
- p1 “Matching offers considerable promise for increasing the reliability and validity of causal inferences...” Definition of “reliability of a causal inference”? I guess this makes sense in the model dependence sense (i.e., the authors aren’t thinking about variation in repeated sampling, but variation over different model specifications). But model space isn’t well defined; all possible subset linear regressions is hardly exhaustive of $g(X_i\beta + t_i\gamma)$.
- p2 “In a sense, our recommendations already constitute current practice...” Indeed. Everyone is doing *something* inferential after some kind of matching, however done. Indeed, some folks do a mix of things after matching (a difference of means for exact matches, and non-parametric inference for inexact matches). There is no shortage of proposals out there; here is another one, parametric analysis after subsetting to the convex hull of where does it line up on bias/variance?
- p30 “...we compute estimates of $\mu_i(1)$ and $\mu_i(0)$ and their uncertainty as usual from the parametric model...” Surely you mean “a parametric model” (the whole rationale for the paper being that inferences are sensitive to the choice of a parametric model, but less so when working with a subset of the data that have good balance).