Political Analysis
Review of "Matching as Nonparametric Preprocessing..."
September 6, 2006

There's a good deal to like about this paper: An important topic, generally clear writing, "clean" examples. It also needs a good bit of work, starting with an effort to overcome its identity crisis. That is, right now it isn't clear to me exactly what the paper's purpose is. Is it an overview of matching methods? A pedagogical/didactic paper? What? My subsequent comments and suggestions should thus all be taken as conditional on resolving that central question.

Irrespective of the answer to that question, a few steps can be taken to make the paper better in general. For one, the current version is repetitive repetitive; there are numerous instances where the author(s) make the same point in two or more places, and in nearly all of these, doing so is neither necessary nor valuable. Additionally, the paper is long – some serious editing could, I think, bring the length down considerably without detracting from its overall message or value.

If (as I'll assume from here on out) the goal of the paper is to provide applied quantitative social scientists with a set of "best practices" for conducting research on causal relationships using observational data, then a number of revisions would make the paper better. First, despite its generally nontechnical tone, the paper still assumes a level of familiarity with matching methods, propensity scores, and the rest of the whole Rubinesque architecture that most applied researchers – even good ones – are unlikely, circa 2006, to possess. Accordingly, a little more detail – including technical detail – on exactly what is going on in such methods would be valuable.

In a similar vein, I'd also suggest that you do a bit more with the two examples. The value of the paper to the end-users of this technique will be much greater if there is some transparency about exactly *how* each of the various steps was taken. Since the MatchIt software is so valuable, some detail about precisely what was done – including the actual information by which things like the degree of balance were selected, etc. – will make the paper a lot more useful.

I would also add a brief discussion about another point. A key aspect of the whole matching business – one that is rightly emphasized throughout the paper – is that it needs to take place prior to any examination of Y, and, in fact, should not include or incorporate information (such as "X_1 has a bigger impact on Y than does X_2) that is gained from Y in the data to be analyzed. This is all well and good if one is working with "virgin" data (that is, where one's priors come from other data), but is potentially problematic when a lot of what we know about the causal relationships among Y, T, and X comes from the data we are analyzing. The field of international relations comes to mind: We have but one realization of international diplomatic history to study, and most (all?) of our information about model specification, functional form, and the like comes from those data. A strict reading of this would suggest that we ought not use what we know to match-and-reanalyze data on (say) international conflicts, since the information we would use to do so it comes from those data. To the extent that at least some readers will be tempted to use the techniques presented in the article to reanalyze (and potentially poke holes in) existing research, the proper place of information from those earlier studies and the best means (if any) to correctly conduct such a reanalysis probably needs some discussion.

A few more picky points:

1. In the Abstract, say "published work," not "articles."
2. The introduction to the Introduction is abrupt, as is the conclusion, and the language in the paper's first 2-3 paragraphs is generally awkward. (I know that's not very specific, but read it out loud and you'll see what I mean). At the same time...
3. ...the Introduction itself seems unnecessarily long.
4. p. 1, bottom: Not everything is a "unified approach," nor is what the paper does anything close to an "inferential framework."
5. p. 4, first full paragraph: The discussion of average quantities of interest could probably do with a bit of expansion. Specifically, it strikes me that some substantial number of social scientists – for example, those doing survey research – are likely to be more interested in population effects than on in-sample effects.
6. p. 5, first full paragraph: A little 2^3 table/typography of the different types of causal effects might be helpful here.
7. p. 8, second full paragraph: This concludes by noting that, absent model dependence, results will be "almost" the same across functional forms. Why the "almost"?
8. p. 9, first and second paragraphs: This entire thing is a bit hyperbolic, and could be scaled down without any loss of effect; *PA* readers can probably understand that 10^V gets big fast in V.
9. Figure 1: On first encounter, Figure 1 is a bit off-putting, since the reader has no idea what the second panel is about. Either (a) separate the two panels, or (b) introduce the figure (and, thus, the example) later in the paper; my own preference would be for (b).
10. p. 15, top: I think the phrase "additional selection" is meant to be italicized.
11. p. 15, second full paragraph: "...control units with X_i greater than $\max(X|T = 1)$ or less than $\min(X|T = 1)$ are discarded."
12. Bottom-of-p. 21 / top-of-p. 22: There are lots of ways to say that Carpenter's moderately-precisely-estimated finding was not in the expected direction, but calling it "statistically significant...in the opposite direction" is not one of them.
13. p. 25: The discussion starting with "We begin with..." can and should be footnoted (since, as a practical matter, it doesn't amount to anything anyway).
14. pp. 25 et seq.: Please refer to "more visible" and "less visible" candidates; "invisible candidates" had me laughing coffee through my nose.
15. p. 27: Figures 3 and 4 are excellent. However, I'm concerned that a casual reader might look at them and think that they are posterior parameter plots or the like. This could lead someone to conclude from Figure 4 that "the analysis without matching found no effect, while that with matching found a small, negative effect that was almost surely bounded away from zero." That would be a wrong interpretation, of course; I wonder if there is some way to incorporate parameter standard errors into those figures to allay this possibility.