PS C236A / Stat C239A
Problem Set 5 - Solutions

1) a) The Mahalanobis distance is defined as:

$$D_m(X_i, X_j) = \left\{ (X_i - X_j)^T S^{-1} (X_i - X_j) \right\}^{\frac{1}{2}}$$

Where S^{-1} is the inverse of the sample covariance matrix of X.

A binary variable with probability of success p has variance $p(1-p)$. A variable with $p = \frac{1}{2}$ therefore has variance of $p(1-p) = \frac{1}{4}$, whereas a variable with p near 0 would have variance near 0. Since we take the inverse of the sample covariance matrix, therefore dividing by the variance, a variable with $p = \frac{1}{2}$ will be given less weight than a variable with p near 0 (or, similarly a variable with p near 1). By FOC, we can show that the variance of a binary variable is greatest when $p = \frac{1}{2}$, so a binary variable with $p = \frac{1}{2}$ will be given less weight than any binary variable with $p \neq \frac{1}{2}$

b) Variables with long tails or extreme outliers tend to have inflated variances, and by the same logic as above, any variable with larger variance will be given relatively less weight.

c) We should be concerned. Outliers and long tails do not make a covariate unimportant, so we may not wish to downweight it relative to other covariates. Binary variables that are very rare may not be of overriding importance, so it may not be wise to give them significantly higher weight than binary variables with p closer to $\frac{1}{2}$. However, if it is a rare binary event, then we might want to treat a difference in outcome as worse than a difference in outcome for a covariate where p is closer to $\frac{1}{2}$. Overall, we should be concerned that Mahalanobis distance exhibits these behaviors for variables for which the theory was not designed.

2) a) The identification strategy in "Fox News Effect" is to tap the variation in Fox news' cable market penetration across localities in order to estimate the effect of access to right-leaning television media on turnout and presidential vote choice. The authors argue that Fox news is introduced by 2000 as-if randomly, *after* conditioning on turnout and Republican presidential vote choice variables in 1996, as well as a number of town-level census demographic variables from 2000 and 1990, and cable system controls. The authors check this selection assumption by regressing Fox media introduction on these controls, in a series of models, some of which include cluster-corrected standard errors and fixed-effect terms for congressional district membership at the town-level.

The model with controls, plus fixed-effects and cluster standard errors reports zero coefficients for the effect of prior presidential vote turnout and Republican choice on Fox entrance in a locality's market. The authors interpret this as recovering conditional independence between treatment and key (i.e. two) selection confounders.

The authors then regress change in Republican presidential vote choice and turnout on Fox news entrance in a regression, including the same controls and fixed-effects specifications that elicited zero coefficients in the above 'selection' regression, interpreting this as the selection on observations assumptions. They find that Fox news had a positive effect on Republican presidential vote.

Weaknesses

The key identifying assumpton is that the included demographic and cable controls are sufficient to recover conditional independence of the potential outcomes of Fox exposure and Fox media presence. The authors provide no extended discussion as to why any or all of these are sufficient or necessary to satisfy this assumption, i.e. there is no theory here. Also, recovering a zero association with the prior outcome is not the same as recovering exchangeability on the potential outcomes conditional on the controls and the model, even if we grant the selection assumption. This itself requires an additional untestable assumption: namely that only the assignment of Fox news, and no other additional factors are relevant in explaining variation in the presidential vote at $t + 1$, beyond those used to model the vote at t.

Let's grant however that the selection assumption holds in general, i.e. these are the correct selection controls for studying the potential outcomes. Unlike in a matching framework where we have an explicit check as to whether conditioning in estimation ensures similarity across Fox and non-Fox markets, the regression evidence suggests there is no linear dependence between the prior outcome and treatment, given the controls and the model. This requires a number of strong modeling assumptions to hold, i.e. correct model for cluster covariance, the fixed-effects model is capturing remaining unit differences (but why not random effects?), stability and unbiasedness in the regression estimates, etc, for the placebo estimates to be informative. This implies that the selection assumption in the regression framework can be quite strong, (and perhaps much stronger than in non-parametric matching), especially since we lose the ability to check for whether we recover similarities on the covariates after conditioning in an separate testing stage.

Finally on a similar point, since a placebo regression was used to identify covariates during a specification search to construct the included model, this changes the interpretation of this as an independent test of a placebo. Given that the search is aimed at specifying a model that recovers a zero here, this test provides no new information that included covariates satisfy the conditional exchageability assumption.

Convincing

The potentially haphazard introduction of Fox news may provide for an opportunity to estimate the causal effect of biased media on political choices and attitudes. However, it is not convincing that the demographic and cable controls contain the comprehensive information used by the Fox organization when deciding where to expand, and thus are sufficient to recover exchangeability. One possible improvement would be to get this covariate information from Fox, and then include that in the selection stage of the design. Having this type of information would seriously improve the plausibility of the result, since it would permit (at least some of) the actual controls used to determine market penetration to be included in the study.

The most convincing piece of evidence is that the authors recover zero effects for Fox market present on the pre-introduction election outcomes. This evidence is suggestive that the model and controls are reducing bias, but not dispositive since these zeros could also be sensitive to model choices and assumptions. Also, see above about how best to interpret this information.

b) This section is very important to establishing the argument that Fox entrance is conditionally random based on the controls and model choices. As mentioned above, my approach would be, at least, to match first and obtain balance on relevent covariates. And then do the regression analysis, after matching, to ensure that the model assumptions do not drive the interpretation that conditional exchangeability holds.

c) The existing data I would want is the same data used by Fox to determine where and when to enter a cable market, as well as additional data that Fox possessed on cable pricing and competition. It would also be great to have similar data for the local cable providers who marketed and made decisions about whether or not to carry Fox above other programming. I would then matching on this data, using GenMatch until balance was obtained before estimating treatment effects. With data limitations, I would still match and obtain balance on whatever controls were available that seem to be plausible predictors of Fox market entrance.

d) Overall, I think the finding is plausible, but not particularly convincing. It's unclear whether the regression model is removing confounding, and the empirical tests of this claim require strong model assumptions. Relaxing these assumptions, and getting more detailed information about the selection process would shore up the strength of the findings.

3) See `HW5_Answers.R` for solutions