

# PS C236A / Stat C239A

## Practice Midterm

### Instructions

This is an ungraded practice exam. The questions are as follows: True/False (Q1) 10%, analytical section (Q2,Q3, Q4) 40%, and the empirical section (Q5 50%). We will not read computer code to find your answer, although computer code necessary to replicate your answers along with its output must be provided with the exam. Ideally, an electronic version of your computer code will be emailed to Danny in one message with the files included as attachments. Please include your last name in the filenames of the attached files.

**Problem 1** An alternative distance metric to the propensity score is Mahalanobis distance. This metric reduces the multidimensional problem of multivariate matching to a unidimensional problem. Although Mahalanobis distance was originally developed for use with multivariate Normal data, we often encounter covariates that are not normally distributed. This problem will explore the implications of these non-normal variables on this distance metric.

- When including a binary variable in a Mahalanobis distance metric, will a binary variable with  $p = 1/2$  or a binary variable with  $p$  near zero be given greater weight by this distance metric? Prove why this is true mathematically.
- How will this distance metric treat covariates with outliers? How about covariates that have long-tailed distributions?
- Should we or shouldn't we be concerned by the behavior of the Mahalanobis distance metric for the covariate distributions described in parts (a) and (b)? Why?

**Problem 2** For this problem, you will compare the critique “The Fox News Effect: Media Bias and Voting”, by Stefano Della Vigna and Ethan Kaplan. The paper can be found here: <http://sekhon.berkeley.edu/causalinf/papers/DellaVignaFoxNews.pdf>. Please write a page or two addressing the following questions:

- Describe and discuss the identification strategy of the paper. What are the weaknesses? What parts do you find convincing?
- Explain the importance of section III.A in the article? Would you do it any differently?
- Perform the following thought experiment: hold the estimation procedure in section III.B constant, and assuming that you have access to all existing data in the US, what data would you include to improve the validity of the estimates? Now do the reverse. Holding the data constant, discuss what parts you would change and add to the estimation procedures to increase confidence in the validity of the results.
- Overall, are you convinced that their conclusions are correct?

**Problem 3** For this problem, you will perform several matching exercises using the “Fox News Effect” data. The unit of observation are towns in the US, and the treatment under study is the availability of Fox News during the 2000 election season. The outcome (`reppresfv2p00m96`) is the change in the Republican presidential vote share between 1996 and 2000. The dataset for this assignment only includes those towns with pre-treatment outcome data, i.e. the change in the Republican presidential vote share between 1988 and 1992 (`reppresfv2p92m88`). The treatment indicator (`foxnews2000`) has been defined as equal to one if the town’s cable system carried the Fox News network before the 2000 election. The dataset includes a set of demographic covariates from the 2000 and 1990 census.

The Fox news data is available here: <http://sekhon.berkeley.edu/causalinf/data/hw6data.RData>. The variables are described in the following file: [http://sekhon.berkeley.edu/causalinf/data/hw6\\_codebook.txt](http://sekhon.berkeley.edu/causalinf/data/hw6_codebook.txt)

For parts (a) - (e) below, be sure to explicitly set seeds to ensure that GenMatch recovers reproducible results, i.e. `set.seed` in general, and in GenMatch `unif.seed`, `int.seed`.

- a. Estimate the causal effect for the treated of a town carrying Fox news on the change in Republican presidential vote share between 1996 and 2000. In doing so, select a set of covariates to condition on, being sure to include higher order terms and interactions you think are appropriate. Also include a propensity score when conditioning, and “orthogonalize” your other covariates using this propensity score. Report your balance statistics before and after matching using `MatchBalance`. Are these effects significant? What is the most interesting summary statistic when comparing change in Republican vote returns? How informative are mean differences? What are the mean differences?
- b. Create a loss function in GenMatch that ensures that the function will not return a matched data set with worse balance on any variable in your `BalanceMatrix` than the balance obtained by matching on *just* your propensity score in part (a) – as judged by eQQ-plots and difference of means. Do this so that this property holds by design – i.e., it holds regardless of the dataset used. (Hint: To do this, you will have to both write a custom loss function and provide GenMatch with starting values for the covariate weights so that it begins with the matched dataset returned from using only the propensity score above.) Match again on your orthogonalized covariates from above using this loss function. Present balance before and after matching using `MatchBalance`.
- c. Now match using the method from part (b) only using demographic covariates. Estimate the “treatment effect” of the introduction of Fox news prior to the 2000 election on the pre-treatment outcome of change in Republican presidential vote share between 1988 and 1992. This is known as a “placebo test”. Can you recover a 0 ATT estimate using only demographic covariates as the conditioning set?
- d. Overall, how do your results differ from those in DellaVigna and Kaplan (2007)? Are your results and their results comparable?
- e. [BONUS QUESTION]. Freed of the constraints in the previous parts, find the best matching method (and possibly post-matching adjustment model) to answer the substantive question at hand. How confident are you that this is an unbiased estimate of the Fox news effect? What do we learn about the effects of media bias from this analysis?