**HOMEWORK # 4. AVDANCED ECONOMETRICS. DUE 30TH NOVEMBER**

This exercise examines the following research question: What is the effect of maternal smoking during pregnancy on infant birth weight and mortality? (see Almond, Chay, and Lee (2005), “The Costs of Low Birth Weight”). Please include a concise summary of your empirical results/answers. We will analyze the following STATA data set:

Data: smoking.dta

The data set comes from the 1989 Linked National Natality-Mortality Detail Files, which are an annual census of births in the U.S., derived from Certificates of Live Birth. Information on subsequent infant death within a year of birth is derived from Death Certificates. This extract consists of all births in Pennsylvania in 1989. The unit of observation is the mother-infant outcome match. There are 139,149 observations on 32 variables. The included variables are:

Variable Description

dbirwt birth weight of infant (in grams)

death 1 if infant died within one-year of birth,0 otherwise

tobacco 1 if the mother smoked during pregnancy, 0 otherwise

dmage mother’s age

dmeduc mother’s educational attainment

mblack 1 if mother is black

mother 1 if neither black nor white

mhispan 1 if mother is Hispanic

dmar 1 if mother is unmarried

foreignb 1 if mother is foreign born

dfage father’s age

dfeduc father’s education

fblack 1 if father is black

fotherr 1 if neither black nor white

fhispan 1 if father is Hispanic

alcohol 1 if mother drank alcohol during pregnancy

drink # of drinks per week

tripre0 1 if no prenatal care visits

tripre1,2,3 1 if 1st prenatal care visit in 1st, 2nd, or 3rd trimester

nprevist total # of prenatal care visits

first 1 if first-born

dlivord birth order

deadkids # previous births where newborn died

disllb months since last birth

preterm 1 if previous birth premature or small for gestational age

pre4000 1 if previously had > 4000 gram newborn

plural 1 if twins or greater birth

phyper 1 if mother had pregnancy-associated hypertension

diabete 1 if mother diabetic,

anemia 1 if mother anemic

1. Under what conditions can one identify the causal effect of maternal smoking by comparing the unadjusted mean difference in birth weight of infants between smoking and non-smoking mothers? Under the assumption that maternal smoking is randomly assigned, estimate its impact on birth weight. Provide evidence for or against the assumption that maternal smoking is randomly assigned.
2. Suppose maternal smoking is randomly assigned *conditional on the observables* determinants of infant birth weight. What does this imply about the relationship between maternal smoking and unobservable determinants of birth weight conditional on the observables? Use a linear regression model to estimate the impact of maternal smoking on birth weight, and report your estimates. Compare the estimate to the one from part (a).
3. Under the assumption of random assignment *conditional on the observables*, what are the sources of misspecification bias in the estimates generated by the linear regression model estimated in part (b)? Now use an approach in the spirit of multivariate matching – i.e., estimate the smoking effects using a flexible functional form for the control variables (e.g., higher order terms and interactions). What are the benefits and drawbacks to this approach?
4. Describe the propensity score approach to the problem of estimating the average causal effect of smoking when the treatment is randomly assigned *conditional on the observables*. How does it reduce the dimensionality problem of multivariate matching?
5. Implement the propensity score approach to the evaluation problem using two methods: 1) control directly for the estimated propensity scores in a regression model; 2) use the estimated propensity score in a subclassification scheme. In doing so, use your own stopping rule – e.g., use the 1% significance level when assessing the balance of the covariates within each block (t-test), and stop the “algorithm” when fail to reject the equality of mean covariates for over 80% of t-tests within a block. Provide empirical evidence on the overlap of the observables of smokers and non-smokers. Estimate the average treatment affect and the average treatment effect on the treated. Interpret the results.

HINT: To use the propensity score in a sub-classification scheme you will use pscore command in STATA.

1. Now use the estimated propensity scores as individual weights (and normalize the weights to one) to estimate: i) the average treatment effect. Compare your estimates to those in part (e) and interpret your findings. What are the benefits and drawbacks of approaches that use the estimated propensity scores as individual weights? Use a graph to provide evidence on the appropriateness of the propensity score weighting estimator – i.e., the sensitivity of the estimated propensity scores. In other words, plot the mean estimated propensity scores (y-axis) against the actual fraction of smokers (x-axis) for 200 equal sized cells of the estimated propensity score (you should produce a graph along the lines of the figure in PAGE 11 of attached PDF Lecture 3). Interpret the results.

**HINT: HERE IS HOW YOU WILL PROCEED FOR GENERATING INDIVIDUAL-LEVEL WEIGHTS. RUN A LOGIT REGRESSION WHERE SMOKING IS THE DEPENDENT VARIABLE AND THERE IS A WHOLE BUNCH OF RIGHT HAND SIDE VARIABLES THAT YOU THINK WILL PREDICT SMOKING BEHAVIOR. AFTER THIS LOGIT REGRESSION, USING THE PREDICT COMMAND generate phat, and (1-phat).**

**logit Smoke x1 x2 x3…..**

**predict phat**

**generate t\_phat=1-phat**

**gen wt = sum (1/phat) where the sum is taken over all mothers who are SMOKERS.**

**replace wt = sum (1/(1-phat)) where sum is taken over all mothers who are NON-SMOKERS.**

**reg BW Smoke [pw=wt],robust**

1. A more general (informative) way to describe the birth weight effects of smoking is to estimate the “nonparametric” conditional mean of birth weight as a function of the estimated propensity score, for smokers and non-smokers. To do this, stratify the smokers into 100 equal-sized cells based on their estimated propensity scores and calculate the mean birth weight and the mean estimated propensity score in each cell. Do the same for the non-smokers. Plot these two conditional mean functions on the same graph, with the mean estimated propensity scores on the x-axis and the mean birth weight on the y-axis. Interpret your findings and relate them to the results in part (e) and (f).
2. Low birth weight births (less than 2,500 grams) are considered particularly undesirable since they comprise a large share of infant deaths. Redo part (f) and (g) using an indicator for low birth weight as the outcome of interest. Interpret your findings.
3. Estimate the impact of maternal smoking on infant death (indicator for death) using the methods in parts (b) and (f). Interpret your findings. From your results, what might you conclude about the relationship between smoking-induced low birth weight and infant death?
4. Concisely and coherently summarize all of your findings. In this summary, describe the estimated effects of maternal smoking on birth weight and infant mortality and whether the causal effect of maternal smoking is credibly identified. State why or why not.