Problem set 1: Empirical Methods for Applied Microeconomics

General instructions. Please work in a group no larger than 3. When you write up your results, please let me know who is in your group. (Only turn in 1 completed homework.). Present your answers in a concise way (typed is highly preferred). Please include relevant Stata output and well-commented do files and ado files for all the exercises (or equivalent in the package of your choice.) Please do NOT include lots of undigested log files. Put the do files in an appendix and make clear reference to the regression output and/or figures.

Problem 1
Measurement error in the March CPS

Download the 2008 March CPS (available at the NBER’s page on [CPS supplement data](https://web.archive.org/web/20080304155704/http://kis.nber.org/cps/cpsdownload.html) (Go to the bottom of the page, March 2008, the box with “A” is the zipped data, that with “P” is the technical documentation for the data, and that with “T” is the data dictionary). The Stata dictionary and do files for reading in the data are at [Programs](https://web.archive.org/web/20080304155704/http://kis.nber.org/cps/cpsdownload.html). You can probably adapt these for your needs. There are also SAS and SPSS files.

We are going to investigate item non-response for wage and salary income earnings, Social Security income, and public assistance income in the CPS, using the insights from Bollinger and Hirsch’s work. Create a data set with all persons 15 or older (the CPS only asks people 15 and older about receipt of income). We are going to explore the allocation/imputations for several variables for the person records.

(i) What share of those 15 and older have an allocated sex or age?

(ii) The wage and salary values like the other income values are collected by a question that says “Did you/name have any income from source “X” last year?” If the respondent says that person received
income from that source, they are asked an amount. It is a bit more complicated for wage and salary income. So, there are 2 questions for the income sources, a yes/no question that is 1 if the person got that income, 2 if they did not, and 0 for out of universe individuals (children 0–14 not asked the question). We are going to look at allocations for 3 income sources for individuals during calendar year 2007: wage and salary, Social Security, and public assistance income.

What share of those 15 or older had the yes/no variables imputed for wage and salary income, for Social Security income, and for public assistance?

What share of those 15 or older had the amount imputed given that they reported that type of income (whether by true report or imputation)?

(iii) Now we are going to see whether this varies by whether the individual responded to the question or it was answered by a proxy. Each household has 1 respondent. You can identify this person by the their line number. The CPS has a hierarchical record structure, where there is a household line, then a family line, and then individual lines. The cpsmar08.do (or .sas) file appended the household and family information to the person records. So, the respondent is the person for whom the respondent number (h_respmn) is the same as their line number (a_lineno). I checked and this identifies the respondent for most records. Does the share of imputed records vary for age and sex for the 15 and older population?

Does it vary for the yes/no questions for receipt of wages and salaries, Social Security, and public assistance?

What about for the levels of the income variables for those who had them?

(iv) What is the relationship between item non-response for demographic variables and yes/no income
variables? What about for the level of income variables?

(v) Model item-nonresponse as a function of some demographics (e.g., race, Hispanic ethnicity, marital status, gender, education, age), month in sample, region, and whether the individual was the household respondent. First create a 0-1 variable for the imputation of the income variables (levels or yes/no). Then, run some linear regressions with robust standard errors. Do you find the same relationships for wage and salary income generally speaking as in the Bollinger and Hirsch articles? Is it different for Social Security income or public assistance income?

Now, use the March Supplement weight (marsupwt). Do the non-response adjusted weights matter for the relationships you’ve found? Do the weights vary with item non-response? (Does it matter what kind of weights you use with robust standard errors?)

Now, we will crudely account for the sample design by “clustering” the standard errors on state (gest-fips). Does this make a difference? Give some reasons why you might think this wouldn’t be sufficient to balance things.

(vi) Run at least one non-linear specification and calculate the marginal effects of gender and whether the individual was a household respondent. Do the estimated marginal effects differ? Consider one interaction (say gender and respondent). Compare the marginal effect from the non-linear specification to that from the linear specification (be careful to account for the main terms). Does the functional form matter?

(vii) Compute inverse propensity score weighted estimate for differences across groups for the natural log of the income variables (as suggested by Bollinger and Hirsch (2006), in their table 2). Compare them to the unadjusted estimates of differences across groups. Are your conclusions different for SS or
public assistance than for wage and salary income?

(viii) Suppose your boss asked what to do about the non-response. What would you do to assess and correct for non-response if you could collect auxiliary data (1/2 page maximum)?

Problem 2

Propensity score weighting.

There is a long standing debate about whether social programs or other interventions can be evaluated without use of data from a randomized control experiment. One of the early entries in this debate was Bob LaLonde’s 1986 AER paper which looked at various non-experimental estimators, using as the comparison experimental estimates. The experiment was the National Supported Work Demonstration, conducted from 1975–77, and Lalonde used nonexperimental data from the CPS and PSID. The NSW looked at the effects of training on disadvantaged groups (female welfare recipients, former drug addicts, parolees, and high school dropouts). Participants had to be unemployed, with little work experience. Obviously, this means these groups were quite different from the general population. Lalonde showed that even using selection adjustments, control groups from the observational data were not able to replicate the experimental findings. Heckman and Hotz (1989) use various tests which exclude the most biased of the estimators from the Lalonde paper.

In this problem, we will use the National Supported Work Demonstration data used in Lalonde (1986). Part of the exercise will be to replicate some of the findings from the Smith and Todd (2005) *Journal of Econometrics* paper which assessed claims in Dehejia and Wahba (1999, 2002) that propensity score matching could be used to replicate closely the experimental estimates. (See the reading list for that and the Lalonde and Dehejia and Wahba papers.) Recall the assumptions under which propensity score matching or inverse propensity score weighting can lead to causal estimates.

Download three data sets from [http://users.nber.org/~rdehejia/nswdata2.html](http://users.nber.org/~rdehejia/nswdata2.html); the original experimental data (either nsw.dta, or the combined nsw_treated.txt and nsw_control.txt), the DW sample from their paper (either nsw_dw.dta or the combination of nswre74_control.txt and nsere74_treated.txt).
and the broadest version of the CPS control group (cps_controls.dta or cps_controls.txt).

For the experimental contrasts, you will either use the original NSW Lalonde sample, or the DW sample. (That is, there are 2 possible experimental contrasts, the original Lalonde one, and the DW one.)

For the attempts to use non-experimental data to replicate the experimental findings, you will use one of the original treatment groups combined with the CPS non-experimental control group.

(i) First, replicate the relevant columns of Smith and Todd table 1 with the Lalonde sample, the DW sample, and the CPS control group. (Note you don’t have the 1979 earnings.)

(ii) Next, test whether the experiment “worked”. Are the means of the $X$s different in the treatment and control groups? Do an overall test for the $X$s being different.

(iii) Based on the time pattern of earnings in the experimental samples in 1974 and 1975, do you see evidence of Ashenfelter’s dip for these participants? (Ashenfelter’s dip is the empirical regularity that those who participate in training or employment programs typically experience a decline in earnings prior to participation. Why would this complicate evaluation using observational data?)

(iv) We will also start by evaluating the simplest possible comparisons. Calculate the basic treatment control difference in the cross-section for 1978 for 3 samples: the full Lalonde sample, the DW sample, and the combination of the Lalonde treatment group and the CPS 1 control group.

Why do you think the CPS 1 control group fails so miserably?

(v) Now we will add some $X$s on the RHS to adjust for differences in observables. Choose some of the controls.
Does adding the $X$s affect the experimental estimates? Should it?

What $X$s help the most with reducing the bias with using the CPS control group? Why might matching on the p-score do better than a linear regression in reducing bias?

(vi) Now, estimate a modified version of the DW propensity score logits (Table 3) for the CPS 1 control group and the Lalonde and DW experimental groups. The dependent variable is a dummy for being from the experimental data, and zero if in the CPS 1 control group. Controls are age, age squared, age cubed, education and education squared, a dummy for being a high school dropout, a dummy for being married, a dummy for being black, a dummy for being Hispanic, real 1974 earnings, real1975 earnings, a dummy for zero earnings in 74 or 75, the interaction of schooling and real earnings in 1974. [Obviously, since we don’t have 74 earnings for the Lalonde sample, we can’t exactly replicate their table 3 column 1.]

Do the coefficients here make sense?

Does it matter if you only use the DW treatment group or use both the DW treatment and control group as the sample for which the experimental data dummy is 1?

(vi) Plot the predicted log odds ratios for the logits for the specifications which include all the experimental observations in each the DW and Lalonde samples. Plot the logodds for the experimental and comparison groups separately. How is the overlap here?

(vii) Now we will look at the bias. Smith and Todd address the bias by comparing the experimental control group (who couldn’t get the treatment) with the CPS and PSID control groups. Differences in these should be zero. Estimate a nearest neighbor p-score impact using the Lalonde sample and choose
a reasonable way of selecting a common support (DW, or what Smith and Todd do is fine). Calculate cross-sectional matching differences. (Analogous to Table 5, row 3). Use the relevant logit generated p-scores for the relevant samples. Implement a nearest neighbor match. Do you think DW are right that the matching takes care of selection?

(viii) Extra credit. Implement a differences in differences matching estimator. (This will subtract a cross-sectional matching estimate of post-RA bias from one of pre-RA bias.) Does this change your view of DW’s claims?