University of California, Santa Barbara **Thomas Heckelei**
Bren School of Env. Science & Management **Winter 2024**

EDS 241
Take Home Final

Due on 2/23/24, 11:59pm
Turn in your R-markdown pdf or html on Canvas in the "Assignments/Exam" section. (including signed cover sheet)

Pledge of honor:
By taking this exam you are pledging to work alone on the exercises. Slack or email me at heckelei@ucsb.edu.edu for any clarifying questions.

Name: Kate Becker

Signature: Kate Becker

Part 1: RCTs, treatment ignorability (selection on observables), propensity scores (15 pts)
This exercise is inspired by Costello et al. 2008 article in science "Can Catch Shares Prevent Fisheries Collapse", which we also discussed in class (lecture 5). "Inspired" means that the data final_fisheries_data.csv are synthetically generated to simplify things for our purposes. It contains the variables on 11135 fisheries (only cross sectional, no time observations): These fisheries were either regulated by an Individual Transferable Quota (ITQ) for all years between 1990 and 2012 or in none of those years. Variables in the dataset are

COLL_SHARE = share of years a fishery is collapsed between 1990 and 2012 (collapse defined as harvest being more than 10% below maximum recorded harvest).
ITQ = dummy variable indicating 'treatment' with an ITQ (equal to 1 if the fishery has been regulated by an ITQ and 0 otherwise).
MET1, MET2, ….MET6 = Dummy variables indicating to which Marine Ecosystem Type (MET) the fishery belongs to (coral reefs, kelp forests, seagrass meadows, open ocean, deep sea, mangrove forests). This type does not change over the relevant time period and does not depend on human influence.
IND_SR = Index of species richness in 1980 with values between 0 and 100 indicating the biodiversity with respect to species in the fishery. Bounds of 0 and 100 are the lowest and highest observed values of species diversity across all fisheries in 1980, respectively.
COMM_VAL = Commercial value of fisheries in 1980 in million US-$

The basic question of interest is "What is the average treatment effect of implementing an ITQ in the time period from 1990 to 2012 on the share of years with a collapse. It is not unlikely that the probability that a fishery is selected for an ITQ depends on the pre-treatment characteristics

given. It is also quite likely that the pre-treatment characteristics have an effect on the share of collapse for each fishery, i.e. our outcome variable of interest.

(a) Compare the distributions of pre-treatment ecosystem characteristics (i.e. MET1, MET2, ,,, MET6) between the treated and the control groups by drawing back to back histograms [2 pts]. Write one sentence discussing the (dis)similarity between the two groups [1pt].
(b) Do a test on mean differences between the treated and control groups for the species richness index (IND_SR) and commercial value (COMM_VAL) variables. Interpret the results (estimated difference and significance) [2 pts] and make a conclusion regarding the similarity between the groups [1pt].
(c) Based on your results from (a) and (b), do you see a problem with just comparing the outcome variable means between treated and untreated fisheries [1 pt]?
(d) Estimate the propensity scores (probability of being treated) using a logit model, assume that all covariates are relevant and should be included in the estimation [0.5 pt]. Draw separate histograms (back to back) of the propensity scores for the treated and the untreated group [0.5 pt]. Comment on the overlap, do you have any concerns? Why/why not? [1]
(e) Use the propensity scores from (c) to estimate the Average Treatment Effect on the Treated (ATT) with a nearest neighbor matching estimator. Interpret the result (just the size of the estimate) [3 pts: 2pt estimate, 1pt interpretation]
(f) Estimate the Average Treatment Effect (ATE) using the weighted least squares on the full sample. Interpret the estimated size and conclude if it is significantly different from zero from a statistical perspective. [3 pts: 2pt estimate, 1pt interpretation]

Part 2 Difference in Difference (DiD) estimation (10 pts + 3 extra credit)

Here we return for a final time to the dataset from Gertler, Martinez, and Rubio-Codina (2012) and use a different way of estimating the effect of the Mexican conditional cash transfer on the value of animal holdings of recipients. We'll use the panel data from assignment 2, where you have both the pre-program and post-program observations. See Template for dataset preparation instructions.

Prepare Data: Load the new data (progresa_pre_1997.csv) and the follow-up data (progresa_post_1999.csv) into R. Note that we created a time denoting variable (with the same name, 'year') in BOTH datasets. Again, you will create a panel dataset by appending the data (i.e. binding the dataset row-wise together creating a single dataset). We want to examine the same outcome variable as before, value of family animal holdings (vani). You will use the full dataset for each estimate. NOTE: you should not change any NAs from the TREATED column in your analysis, as we expect that spillover was likely in this program. NAs will be excluded from your calculations/estimations

(a) DiD Estimator, ATE (5 pts: 3 pts estimate, 2 pts interpretation): Calculate the DiD estimator of the treatment effect (ATE) of the program on the value of animal holdings (vani) "manually" i.e. based on group mean values without running a regression. Report and interpret the result (Note: no significance test or standard errors is possible, so you do not need to report these values).

(b) Difference in Difference using OLS (5 pts): Now set up an OLS-regression using group mean values to estimate the same ATE. Interpret the estimated treatment effect [3 pts]. Also interpret the coefficients on the time dummy and the group dummy variable (see interpretation done in class in lecture 9) [2 pts].

Hints: You will need to create a new dataframe with a variety of dummy variables to do this. The R example provided with the DiD module (and/or the excel file) should help.

(c) Extra Credit: ATE with OLS using full dataset (3 pts: 2 estimate, 1 interpretation): Estimate the ATE with an OLS-regression based on the original units as observations (i.e. not with group mean values, you will need to use the entire dataset). Even though the specification is the same as in the regression with the group mean values above, you'll need to create new indicator variables for the treatment group and the post treatment time period as well as their interaction term. Verify that you get the same result as above. Now report also on the precision of the estimation and test whether the estimated coefficient is different from zero.