**1. Regression discontinuity**

a.

The running variable is income, and the cutoff is 20.000 $.

b.

Sorting means that for whatever reason there is a discontinuity in the running variable’s distribution at the cutoff. In practical terms, it means that observation accumulate either just before or just after the cutoff for endogenous reasons.

The first step in evaluating the presence of sorting is analyzing the running variable’s distribution graphically, most of the times sorting can be seen by the bare eye. The next step is performing data-driven tests whose objective is to reject the hypothesis that the distribution is the same. In particular, I would perform McCrary and Cattaneo et al. tests and compare their results.

c.

Barrecca et al. suggest dropping the units in the vicinity of the cutoff and to re-estimate the model using only the units in the neighborhood. This is called a ‘donut’ RDD.

d.

Since we want to find whether a certain treatment had a causal effect on the outcome variable, we would like the observation’s potential outcomes, as well as other variables to follow a smooth pattern instead of presenting a jump in the cutoff. In other words, we would like the observations on both sides of the cutoff to be very similar in every variable except on the treatment, so that we can determine that the treatment indeed caused the observed difference in the outcome variable. That is what we mean by smoothness assumption.

In this application, the smoothness assumption implies that people whose income is 20.000 USD and those whose income is 19.999 USD are very similar on average in terms of education, employment status, and every other which may plausibly affect child health outcomes. If we find that, we can assume that their potential outcomes were similar, and thus that any discontinuity was caused by the policy

e.

The first auxiliary test is called a balance test, and it is performed for every covariate. It simply consists in running the RDD with the covariate as dependent variable. If the model shows a discontinuity for such covariate in the cut-off, there is strong evidence of non-smoothness.

The second test is called a placebo test. When performing this test, the researcher uses the RDD to evaluate if there is a discontinuity at fake cut-offs. In this example, it would mean searching for a discontinuity with 5.000, 10.000, 15.000, 25.000, 30.000, etc as cutoffs. If those models yielded evidence of discontinuity, that means potential outcomes were most likely not smooth at the real cutoff.

**2.Instrumental variables**

a.

Bias of 2SLS estimators will be centered on the bias of OLS estimator if the instrument used is weak (not strongly correlated with the endogenous variable in the first stage).

b.

By showing that the first stage of our instrument is strong. We may do so with the First stage F statistic. We should also make a case for it (convince others that the instrument is reasonable), by appealing to previous knowledge on the matter.

c.

\*Instrument strength: There is a strong correlation between the instrument and the endogenous variable. In this particular case there is a strong correlation between winning the lottery and participating in the program.

\*Exclusion restriction: There is no correlation between the instrument and other unobserved factors relevant to the dependent variable. In this case that means that winning the lottery is independent of unobservable relevant variables such as interest of the parents in the child’s wellbeing.

d.

Those terms refer to how the instrument affects the treatment status. Heterogeneous treatment effect estimators only affect the complier group.

The complier group is composed of people whose treatment status was modified by the treatment status. In this case, it refers to people who enroll in the program because they won the lottery, but would have not enrolled otherwise. The defiers are the exact opposite: this group comprises people who enroll having lost the lottery, but that would not enroll if they won the lottery. The never takers are people who will never enroll regardless on whether they have won the lottery, while the always takers are those who always participate in the program, regardless of having won the lottery.

e.

There are not, because participation in the program is conditional on having won the lottery, so no one could participate if the don’t get selected by the lottery.

f.

Mathematically, both the compliers and the defiers do, but since a common assumption is that there are no defiers, we conclude that only compliers do.

g.

The parameter estimated is a Local Average Treatment Effect (LATE)

h.

Because simply comparing people on the program with people outside of it provides a biased estimator of the effect of the treatment in the program. In this particular case the bias would be caused by omitted relevant variables that cause both the enrollment in the program and the health outcomes such as parents’ commitment on their children’s health.

**3. Difference-in-differences**

a.

\*We need regressors to be strictly exogenous conditional on the unobserved fixed effect

\*We need regressors to vary over time at least on one period, and for them not to be perfectly collinear.

b.

No, the estimator yielded will be biased. It will attenuate the estimator and may even change its sign. The reason behind this is that a non-constant ATT creates heterogeneity bias. Time varying treatment effect generate cross-group heterogeneity because of differing post-treatment periods. The treatment effect has a break (non linearity) in trend, and thus its estimator will be biased

**4.** **Difference-in-differences**

a.

Variance Weighted Average Treatment effect on the Treated (VWATT) : This is the parameter we are estimating (instead of the normal ATT). It is weighting each individual set of ATTs with each weight depending on group size and variance.

Variance Weighted Common Trends (VWCT) : It is a sort of parallel trend bias term. It does not necessarily need for every trend to be equal and is thus a weaker assumption than a pure parallel trend assumption. The weights can be adjusted to correct for unequal trends and eliminate the bias.

Heterogeneity bias (: This term introduces bias when ATT varies across units and/or when it changes in time (Dynamic treatment effects). It is an attenuation bias and will cause an underestimation of the treatment effect.

b.

Considering there are 4 timing groups and a control (untreated) group, there should be 16 distinct 2x2 estimates

c.

I will calculate the variance of treatment

T0 = 1980, T1 = 1985, T2 = 1990, T3 = 1995, T4 = 2000, T5 = 2005

Var Group 1: 0.16. Var Group 2 = 0.24. Var Group 3 = 0.24. Var Group 4 = 0.16

d.

The largest weight will be assigned to groups treated on the middle of the panel and when the difference between treatment variance is close to 0.5. That is because of the way the weights (both treated on treated, and treated vs untreated) are built. They are built to up weight groups that spend nearly half of the time with treated status.

e.

If treatment occurs first to units with the largest ATTS, then the estimator underestimates the sample-weighted ATT if the first treatment occurs early enough, and overestimates it if it happens late enough. They consequently suggest to scatter the weights according to each group’s sample share to avoid that.

f.

It will yield the ATT under the assumptions of parallel trends and homogeneous treatment effect.

g.

Having dynamic treatment, and/or having difference in treatment timing will bias the estimator. It will attenuate the estimator and may even change its sign. The reason behind this is that a non-constant ATT creates heterogeneity bias. Time varying treatment effect generate cross-group heterogeneity because of differing post-treatment periods. The treatment effect has a break (non-linearity) in trend, and thus its estimator will be biased

h.

The main parameter of interest is the Group-Time ATT

i.

\* IID sampling from panel data

\* Parallel trends (conditioning on observables)

\* Irreversible treatment (A unit cannot pass from treated to untreated

\*Common support (The propensity score is defined during the whole domain)

j.

Number of groups\*Number of periods = 4\*25 =100

k.

g denotes groups and t denotes periods of time. denotes a particular group and is binary, equalling one if the group is treated on period t. C is a binary indicating a control group that equals one if has never been treated. is the propensity score calculated as . In other words, it is estimated based on group covariates using probit or logit. Y denotes the outcomes.

l.

They propose to use the “not yet treated” as a control group until evry group becomes treated.

m.

Because of the need to use the ‘not yet treated’ groups as control, it will not be possible to find the last group’s group-time ATT, since there will be no control group to make the comparison.

**5. Synthetic control**

a.

It does not require access to post-treatment outcomes in the“design” phase of the study, so it can be more intuitively used for policy design. It makes explicit the contribution of each comparison unit to the counterfactual. It formalizes the way comparison units are chosen, because it has direct effect on the inference and finally, it precludes extrapolation because the counterfactual forms a convex hull.

b.

The estimator, called a Dynamic ATT, is the simple difference between the actual outcome, and the predicted outcome (based on the covariates of the chosen combination of the controls), for every moment in time. The predicted outcome is a function of the outcomes of the ‘similar units’ in the sample, weighted for their ‘alikeness’ to the treatment unit.

c.

The minimization problem is to minimize the squared sum of prediction error (MSPE) for the pre -treatment error, by changing the weights allocated to each of the units in the sample (except the treatment unit). The maximum weight that could be assigned is 1, and in such a case it would mean that there is an observation in the untreated group that was identical to the treated unit, and thus constitutes the perfect counterfactual.

d.

The method recommended is to use Fischer’s sharp null with an statistic equal to the ratio between RMSPE after and before the treatment. This method requires to iteratively apply the synthetic algorithm to each unit in the donor pool and obtain a distribution of placebo effects to compare the gap (RMSPE).