Exam 4137 – Spring 2023 - postponed

IMPORTANT: Always motivate your answers. Show knowledge and understanding of the concepts taught in the course. Make sure not only to repeat material from lectures and curriculum, but engage with the empirical examples and show that you understand what it means in the particular setting. Your answers should be as short as possible and as long as necessary. Each lettered subtask carries 10 marks, for a total of 140 points.

---

1. True or false:

   (a) Homoskedasticity is required for Ordinary Least Squares to give unbiased estimates.

   FALSE. Homoskedasticity is the assumption that the variance of the error terms is constant and unrelated to regressors. It is required for consistent estimates of the **standard errors of OLS coefficients**, but not for unbiased estimates of the coefficients themselves..

   (b) In a randomized controlled trial with noncompliance, all we can identify is the intention to treat-effect (ITT).

   FALSE. We can use instrumental variables to estimate Local Average Treatment Effects, which are average treatment effects for individuals who comply with treatment assignment and take treatment when assigned the treatment group and not otherwise.

   (c) The standard error of a maximum likelihood estimator is larger if the curvature of the log likelihood function at the optimum is smaller.

   TRUE. The variance of a maximum likelihood estimate is the inverse of the information matrix, which is the average of the negative of the curvature of the log likelihood function. Intuitively, when the curvature is large and negative at the optimum, the log likelihood reduces a lot as we move away from the optimal point, it is easier to exactly pinpoint the optimum an we get a precise estimate.

   (d) The marginal effect at the average from a probit or logit model is often very close to the coefficient from a linear regression model

   FALSE. The coefficients from linear regression models are typically closer to the average marginal effect than the marginal effect at the average. Students may explain the difference between APE's and PEA's.

   (e) In a randomized controlled trial with onesided noncompliance, we can estimate the average treatment effect for the treated or the average treatment effect for the untretead

   TRUE. With onesided noncompliance, either all treated individuals are compliers (if people with Z=0 cannot get treatment) or all untreteated individuals are compliers (if people with Z=1 always take the treatment). In the first case, we can estimate the ATT, in the second we can estimate the ATUT.

2. The Norwegian teacher's union and various policymakers regularly advocate for smaller classes in order to improve student learning. Starting in the fall of 2018[1], a new norm was introduced that stated that there should be at most 15 students per teachers in grade 1-4 and at most 20 students per teacher in grades 5-7. Imagine that you work in the Ministry of Education and are asked to predict how such a norm would affect student test scores. You have data from the year 2017, and run the following regression:

$$y_g = \alpha + \beta CS_g + \varepsilon_g$$

Where y_g are the average test scores of all students in school g at national tests taken at the end of the year and CS_g is the average class size of school g. In this exercise, you do **not** need to discuss the timing of treatment or the scale of test scores.

(a) What assumption do you need to make for β to be interpreted as a causal estimate of the effect of class size on test scores? In the likely scenario that this assumption does not hold, what is the likely direction of the bias? Argue why.

Randomization, CS_g uncorrelated to the error term ε_g. Students may also present this in terms of potential outcomes, but notice that we have not explicitly covered potential outcomes with multivalued treatments such as class size so this is not necesssary.
For the direction of the bias, it does not matter how the students conclude, only how they argue. Omitted variable bias is equal to the correlation between the outcome and the omitted variable times the correlation between the variable of interest and the omitted variable. We can think about the omitted variable as being for instance an urban dummy. If urban schools are likely to have better scoring students (independently of class size) and also large class size, both correlations are positive and omitted variable bias is positive - thus the estimate we get of β is too large.

After the introduction of the new norm, you get access to data from 2018, the first year when the reform was in place. You are tasked with evaluating whether the policy increased test scores. You run the following regression for the years (t) 2017 and 2018

$$y_{gt} = \theta_g + \gamma_t + \eta T_g \times 1[t \geq 2018] + \mu_{gt}$$

Where $T_g = 1[CS_{g,2017} > \bar{CS}]$ is a dummy for whether average class size in school g in 2017 was larger than the maximum class size $\bar{CS}$. $1[]$ is the indicator function. The outcome is either average class size or average test scores. θ_g are school fixed effects while γ_t are year fixed effect. You obtain the following results:

[1] 2018 was a phase-in year with a slightly higher norm, but in this exercise you may disregard this detail.

Table 1: Regression results

Outcome	Class size	Test scores
Coefficient on T_g	-2.28	.159
std. error	(0.47)	(.062)

(b) What identifying assumptions do you need to make in order to give the two parameters in the table a causal interpretation? What interpretation is that?

This is essentially a difference in differences application, estimated with two-way fixed effects, and we need parallel trends between control- and treatment schools. Parallel trends mean that in the absence of the teacher norm, schools with lower average class size than the norm and schools with higher average class size than the norm would have followed the same trend in average test scores. If this is the case, the coefficients in table 1 can be interpreted as (a weighted average of) ATT's: The average treatment effects of the policy on respectively class size and test scores for schools with higher than norm average class sizes in 2017.

(c) What additional assumptions do you need to make to use the exposure to the teacher norm as an instrumental variable in this setting? Are they plausible? Are they testable?

We need exclusion: The teacher norm cannot affect test scores other than through class size. This is not testable, but students could briefly discuss possible violations.
We also need relevance: The policy must affect class size. This is testable, and we see from the table that $F = t^2 = (-2.28/0.47)^2 = 23.5 > 10$, well above the rule of thumb of an F-statistic of 10.
Finally we need monotonicity: Having average class size above the norm in 2017 cannot induce any schools to increase class size. Note that we have only briefly covered potential outcomes with multivalued treatments, so a text explanation will be sufficient.

(d) Calculate and interpret the instrumental variable estimate of the effect of class size on student test scores.

$IV = \frac{RF}{FS} = \frac{0.159}{-2.28} \approx -0.07$. Each increase in class size by 1 student reduces average test scores by 0.07 among schools that reduce class size in response to the teacher norm, but wouldn't otherwise.

(e) Would you account for heteroskedasticity or clustering in your calculation of standard errors in this setting? Why or why not?

In this case it is likely that the homoskedasticity assumption is violated, either because schools with different average class size have different variance of the error term, or because error terms are correlated within schools across years. Thus, cluster-robust inference may be appropriate as long as the number of schools is large enough, which is also heteroskedasticity-robust. Students may also mention the Moulton factor.

You eventually dig up more data to cover the years 2010 to 2020, rerun the reduced form regression above and produce the plot in figure 1.

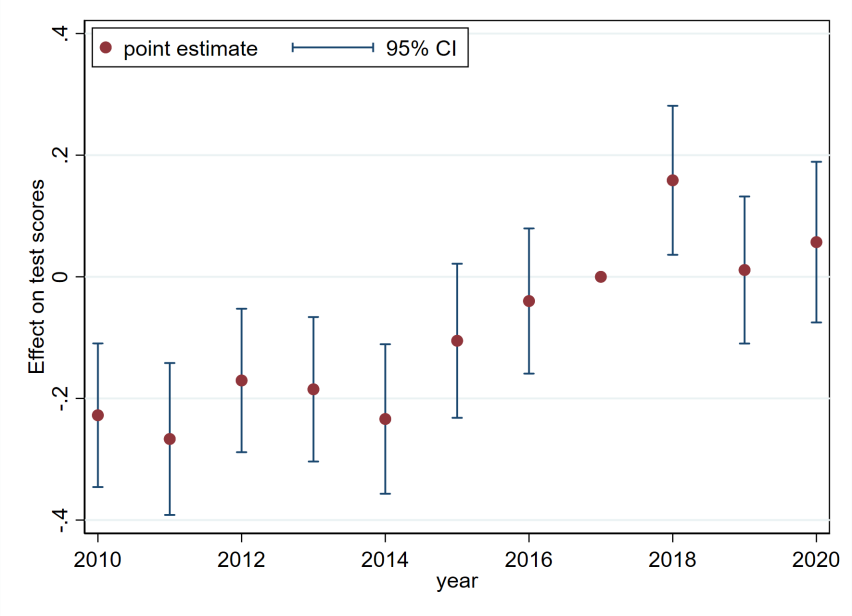


Figure 1: Reduced form Difference- in differences estimates over time

(f) Do these support the common trend assumption? Why or why not? Provide an alternative explanation for the positive effect on test scores in 2018.

Common support is fundamentally untestable, so a test of parallell trends cannot prove or disprove the assumption. It can, however, lend support to the assumption: If the trends in the treated and control group are similar in the years leading up to 2018, it is more plausible that they would have been similar also after 2018 in the absence of the policy. This figure does not support the common trends assumptions, as treated schools (=schools with above norm average class sizes in 2017) are clearly on a steeper trend than schools that had smaller class sizes. There could for instance be increasing test scores in urban areas in the 2010's compared to rural areas, which would vilolate common trends. The effect for 2018 looks more like a continuation of this trend, and we would worry about violations of common trend.

3. During the COVID pandemic, governments struggled with implementing appropriate countermeasures under limited information. The paper *"What the COVID-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan"*[2] evaluates the effect of school closures on child wellbeing and lifestyles. They exploit the sudden and unprecedented request for closure of all schools by Prime Minister Shinto Abe in Japan on March 2, 2020, while kindergartens and preschools remained open. They estimate the following equation

$$Y_i = \alpha_0 + \alpha 1[m_i \geq 90] + \alpha_L(m_i - 90) + \alpha_H 1[m_i \geq 90](m_i - 90) + \varepsilon_i$$

Where Y_i is an outcome - either a dummy for child i not attending school in March 2020 or a measure of child wellbeing - and m_i is a child's age in months at the time of the survey. In Japan, children start school on April 2 if they have turned 6 years of age by that day, so that at the time of the survey the following year, children less than 90 months of age would not have started school in March 2020 and children 90 months or above had started school by that time.

(a) What assumptions does the authors need to make? Can you suggest a potential violation?

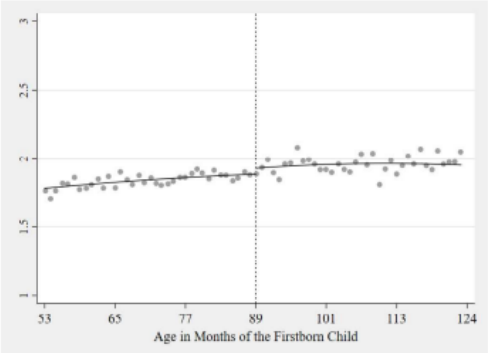
The core identifying assumption for a regression discontinuity approach is contininuity: The potential outcomes are continuous in the running variable at the cutoff. $\lim_{m\downarrow 90} E(Y_g(m)) = \lim_{m\uparrow 90} E(Y_g(m))$ for both potential outcomes $g = 0, 1$. In this case, expected outcomes given that you attended school in march 2020 cannot exhibit a jump at 90 months, and the same for the expected outcome given that you are not in school.
This is violated if anything else changes as the cutoff. In this case, 90 months of age also marks the cohort change: Children above the cutoff are in first grade, children below have not yet started. Moreover, children that are 90 months are the youngest in their cohort, children that are 89 months old are the oldest in theirs. If having started school or being young or old in your cohort affects child well-being, we might have a violation of the continuity assumption.

(b) In Figure B1, the authors provide support for the identifying assumption. What is the intuition behind this figure? Does it indeed support the identifying assumption? What kind of variable would you ideally like to use for this exercise?

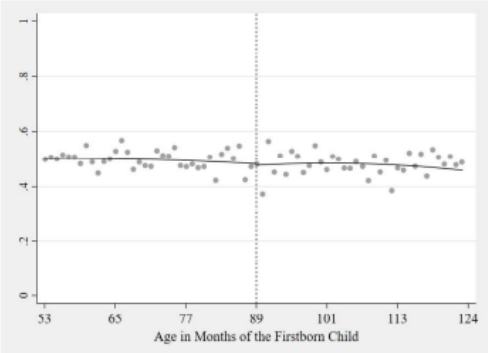
[2] Reo Takaku and Izumi Yokoyama, Journal of Public Economics Volume 195, March 2021.

Figure B1: Check for Continuity of Potential Covariates

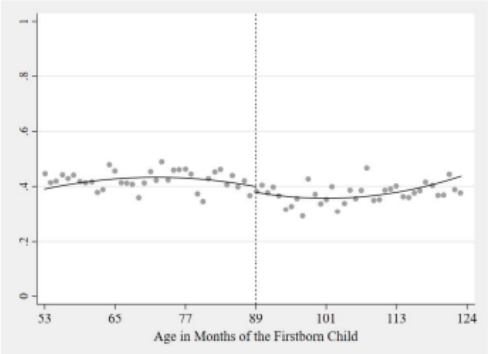
(a) Number of Children



(b) Firstborn Child Is a Girl



(c) 4-Year College Graduates



(d) Support from Grandparents Available (As of Feb.)

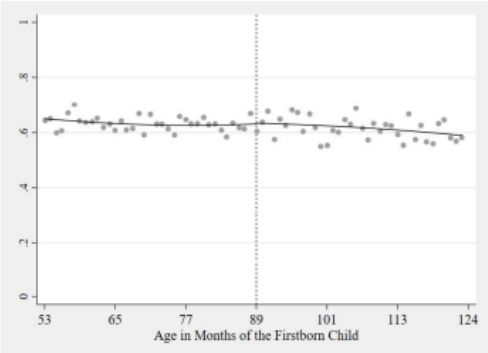


Figure 2: Continuity checks in Takaky and Yokoyama (2021)

This is a continuity test, aimed to investigate whether children to the right and the left of the cutoff are different on background covariates. If they are, we would worry that these covariates could also affect outcomes or correlate with things that affect outcomes, independently of school attendance. The tests seem very smooth at the cutoff, supporting the identifying assumption. For this kind of test, we ideally want covariates that strongly affect the outcome - child well-being and nutrition. The variables used in the figures are perhaps not the best examples. If one would for instance want to look at the potential violation suggested above, one could instead look at things like self-reported well-being or daily nutrition measured before Covid, but after school start.

(c) The main results of the paper is summarized in table 2. What assumptions do the authors need to make for the estimate in the "fuzzy" column?

Table 2: Regression discontinuity results from Takaku and Yokoyama (2021)

Outcome	Sharp	Fuzzy
My child gained weight	0.092	0.154
	(0.022)	(0.042)
My child did not go	0.623	
to school/preschool	(0.027)	

Fuzzy RD is just an instrumental variable approach. Therefore, we need exclusion, relevance and monotonicity in addition to the continuity assumption above. Exclusion imply that crossing the threshold only affect outcomes through school attendance. Monotonicity implies that crossing the threshold does not shift any children into school. Relevance imply that there is a first stage: Crossing the threshold shifts at least some individuals out of school. The latter is testable, and we see that the first stage estimate in row 2, column 1 is strong and significant - 62.3 percent of the sample are compliers.

(d) Interpret the estimates in the "fuzzy" column. For what part of the population does this represent average causal effects?

The school closings in Japan in March 2020 caused 15% of children who did not attend school in march 2020 to gain weight, based on mother's self reporting. This is an average effect for a) compliers who would attend school if they were just young enough to not have started first grade but would have started if they were just old enough and b) for children of age close to the 90-months threshold, meaning that they would be around 84 months at the time of the school closures.