Please collect the answers to the questions below (including any tables and figures) in a `.pdf` and submit it together with the `.do` files that generated any tables and figures on Athena under the corresponding folder. If you write with pen/pencil, make sure to have clear photographs for submission. The deadline is 16th of May, 16:00.
Please remember to lay out your results as clearly as possible, and to comment on your code in a way that makes it easily accessible to others.

Question 1. (Exam 2023: A dynamic panel data process.)
Consider a setting with two cohorts: one treated at $g < \infty$, and a control cohort ($g = \infty$). Units also differ in terms of a binary characteristic $X_i \in \{0, 1\}$, with

$$0 < \Pr(X_i = 1|G_i = g) < \Pr(X_i = 1|G_i = \infty) < 1.$$

Imagine the following data-generating process:

$$Y_{it} = \phi(X_i) + \rho Y_{it-1} + D_i P_t \lambda + \varepsilon_{it},$$

for all $t = 2, ..., T$, where $\phi : \{0, 1\} \to \mathbb{R}$, $\rho \in \mathbb{R}$, $D_i = 1[G_i = g]$, $P_t = 1[t \geq g]$, $\varepsilon_{it} \sim F$ with $\mathbb{E}[\varepsilon_{it}] = 0$, and $Y_{i1} \sim N(0, 1)$.

1. How would you interpret the function $\phi(X_i)$? What about λ?

2. Set up a dynamic DID estimator. Under what assumption(s) about the parameters/functions does it identify λ?

3. Imagine these assumptions are <u>not</u> satisfied. Set up an alternative estimator that identifies λ. What class of estimator is it? What weaker assumptions does it need to satisfy?

4. Imagine you get data from an additional cohort with $G_i = g + 1$ (i.e. it is treated one period after the original treated cohort), with $0 < \Pr(X_i = 1|G_i = g + 1) < 1$. What would be the benefits of this additional data? What kind of changes to your estimation strategy would it require?

Question 2. (Exam 2022: College scholarships as DID or RDD.)
Imagine a cohort of college students that may receive a scholarship $D_{it} \in \{0, 1\}$ in the first two years of attending college if they score above the time-varying cutoff $c_t \sim U[c_L, c_H]$ at the beginning of each year. If they are awarded the scholarship in the first year, they automatically receive it in the second year as well. If they are *not* awarded the scholarship in the first year, they can try again in the second year. Overall, a student may receive zero, one, or two scholarships and there is no attrition. Student $i = 1, ..., N$ scores $R_{it} \in [0, \bar{R}]$ in year t with $0 < c_L < c_H < \bar{R}$ and assume that $F(r) = \Pr(R_{it} \leq r)$ is positive for all $r \in [0, \bar{R}]$. We are interested in their potential academic performance in each year $Y_{it}(d)$ which depends on whether they received the scholarship at the beginning of that year or not $d \in \{0, 1\}$.

1. Consider the panel of students observed across these two years. What is the treatment structure? Write down a matrix describing the grouped treatment structure. Define an average treatment effect on the treated τ. Consider exploiting this setting for a DID design. Under what assumptions is τ identified?

2. Define a switching estimator τ_S in the style of de Chaisemartin and d'Haultfoeuille (2020). Under what assumptions is τ_S identified? Compare its target parameter to τ. How would you estimate this new parameter?

3. Consider now the fact that the scholarship is awarded discontinuously on the basis of the cutoff c_t. Under what assumptions is $\tau_t(c)$, the treatment effect of students with $R_{it} = c$, identified? Define this treatment effect in terms of potential outcomes. How does it compare to τ_S? Write down an estimator for $\tau_t(c)$.

4. Imagine exploiting both the panel structure and the discontinuity simultaneously. What parameter would this procedure identify? How do the assumptions differ under which this parameter is identified from before? Write down an estimator of this parameter.

Question 3. (Exam 2022: Unbalanced difference-in-differences, with coding example.) Consider the following treatment structure:

$$\begin{bmatrix} \cdot & \cdot & 0 & 1 \\ \cdot & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

where the dot denotes missing data.

1. Consider the following two-way fixed effects (TWFE) model:

$$Y_{it} = \alpha_i + \gamma_t + D_{it}\beta + \varepsilon_{it}.$$

with parameters α_i for all i, γ_t for all t, and β. Which parameters are identified? Why?

2. Write down an equation for a dynamic difference-in-difference (DID) estimator. What causal effect(s) does this estimator identify? Under what assumptions?

3. How does one generally test the assumption underlying a dynamic DID estimator? How does this apply to the treatment structure in this question?

4. Consider now the possibility of heterogeneous treatment effects. Imagine all treatment effects $\mathbb{E}[\tau_{gt}] = 1$, except the one on the bottom-right of the treatment structure, $\mathbb{E}[\tau_{14}]$. What is the largest value this treatment effect could take on without changing the sign of the TWFE estimate of the treatment effect?

5. Generate a sample of data corresponding to this treatment matrix. Specifically, assume that $T = 4$, $G_i \in \{2, 3, 4\}$, and observations before $g - 1$ are missing for each group. Further assume that

$$Y_{it} = \alpha_i + \gamma_t + D_{it}\tau_{it} + \varepsilon_{it}.$$

and that each of the groups have $N_g = 100$ units; $\alpha_i \sim U[0, 1]$; $\gamma_t = 0$ for all t; and $\varepsilon_{it} \sim \mathcal{N}(0, 1)$. Assume that the true treatment effects $\tau_{it} = Y_{it}(g) - Y_{it}(\infty)$ are zero before the treatment (i.e. no anticipation); in the period of treatment, we have $\tau_{it} = g/2 + u_{it}$ with $u_{it} \sim \mathcal{N}(0, 0.1)$, meaning that later treated units experience a larger initial treatment effect. For subsequent periods, i.e. event times 1, 2, etc, we have $\tau_{it} = \tau_{it-1} - t/5 + \xi_{it}$ with $\xi_{it} \sim \mathcal{N}(0, 1)$, meaning that treatment effects drop off over time.

6. Use `panelview` or a similar command to display the treatment matrix based on your simulated data. Comment on the potential of using a dCdH switching estimator here. Would it work, and if so, how?

7. Make a figure of the average outcomes of each group over time in a single plot.

8. Estimate a TWFE regression and compare the coefficient estimate on D_{it} to the true ATT, defined as

$$\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}D_{it}}$$

9. Estimate a dynamic DID with separate estimates for the period of treatment and the period right after treatment. How does the estimate for the period of treatment compare to the ATT of the first period, defined as:

$$\frac{\sum_{i=1}^{N}\tau_{ig}}{\sum_{i=1}^{N}\mathbf{1}[G_i = t]}$$

   Comment on the difficulties of estimating causal effects in this setting.