Exam 4137 – Spring 2025, 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 points, for a total of 140.

---

1. (20 points) Consider an employment training program (T) designed to improve earnings (Y) among unemployed individuals, and a subsidy (Z) to the price of the training program. Define Y_{dz} as potential earnings with treatment exogenously set to d, and instrument to z, T_z as potential training with the instrument set to z and Y_d as potential earnings with treatment set to d.

   (a) Explain the following estimands in light of the example

      i. $\mathbb{E}(Y_1 - Y_0|T = 0)$

      ii. $\mathbb{E}(Y_1 - Y_0|T_1 > T_0)$

      i. This is the average treatment effect for the untreated - the ATUT. In the example, this is the average difference between earnings with and without the training program for individuals who do not take the training program.

      ii. This is the local average treatment effect for compliers to the instrument - the LATE. In the example, this is the average difference in earnings with and without the training program for individuals who take the program when it is subsidized, but wouldn't otherwise.

   (b) Explain what the following assumptions mean. Are they reasonable in the current example?

      iii. $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, T_0, T_1) \perp Z$

      iv. $\mathbb{E}(Y_1 - Y_0|T = 1) = \mathbb{E}(Y_1 - Y_0|T = 0)$

      i. This is the instrument randomization assumption, which states that all potential outcomes and potential treatments are independent of the realization of the instrument. In this case, it implies that individuals who are assigned the subsidy have statistically identical earnings and treatment choices for all potential realizations of the instrument and treatment. In this setting this is reasonable if the subsidy was randomized - otherwise it depends on how the subsidy is allocated.

      ii. This assumption states that there is no selection on gains: The average difference between treated and untreated potential outcomes is the same for treated and untreated individuals. This is probably unlikely in this setting - we would expect that rational individuals who have larger treatment effects are more likely to take treatment than those with smaller, implying $\mathbb{E}(Y_1 - Y_0|T = 1) > \mathbb{E}(Y_1 - Y_0|T = 0)$.

2. (80 points) A crucial question for public policy is to what extent taxes depress labor supply, but comparing individuals with high and low tax rates are unlikely to recover causal estimates. Recently, the Ministry of Finance proposed to randomly assign lower taxes to 100,000 individuals 30 years or younger, for a period of 3-5 years. In this exercise, you may ignore the extensive margin of labor supply and think only of intensive margin (earnings).

   (a) If the program is randomly assigned, what parameter does a regression of log earnings ($\log(Y)$) on program participation (D) identify? Discuss how you could use this setting to estimate an elasticity of earnings with respect to taxes.

   Under randomization ($Y_0, Y_1) \perp D$), the simple comparison of means of log earnings identifies the ATE for young individuals - the average difference in earnings between young individuals in the low-tax regime and high tax regime. Because the outcome is in logs, the treatment effect is roughly interpretable as percentage changes in earnings. When the treatment is randomized, there can be neither selection on levels nor gains.
   If we want to estimate an elasticity of earnings with respect to taxes, we would think of this estimate as a reduced form effect. If we instead wanted to get the effects of lowering the tax rate, we could instead instrument for the log of the (net of tax) rate, log(1-t), using the treatment dummy, under additional assumptions such as exclusion.

   Unfortunately, the proposal was heavily debated and might be abandoned. Imagine instead that the lower-tax regime is introduced at last minute to all individuals aged 30 or below starting in 2026, and you have access to panel data on individuals' earnings and age over time before and after the introduction of the reform.

   (b) You want to estimate the effect of the low-tax policy on log earnings using Differences in Differences. What regression model would you run?

   $$\log(Y_{it}) = \alpha + \theta_t + \beta_{age_{it}} + \gamma \mathbb{1}(age_{it} \leq 30, t \geq 2026) + \varepsilon_{it}$$

   , where θ_t are year fixed effects, β controls for any level differences between young and old individuals and γ, the parameter of interest, captures the differences over time in young and old individuals before and after the reform. Students are free to change age dummies to a single young versus old dummy, and could also consider adding individuals fixed effects if they make and argument for why - for instance to improve precision by reducing residual variance. It is also perfectly fine to add a POST dummy ($\mathbb{1}(t \geq 2026)$) in place of year fixed effects.

   (c) What assumption do you need to make for the regression model to estimate a causal effect? When treatment effects are heterogeneous, what causal effect is that?

As all Difference in Differences models, this relies on common trends: $\mathbb{E}(Y_1^0 - Y_0^0 \mid age \leq 30) = \mathbb{E}(Y_1^0 - Y_1^0)$, where Y_t^0 are potential log earnings without the tax reform in time $t = 0, 1$ for pre/post. The difference between log earnings of individuals age 30 or less in 2026 and later vs 2025 and before should be the same as the same difference for above-30-year olds. To get full marks, students should note that this implies that the *relative* change in the earnings of the two groups should be parallel, because the outcome variable is in logs. Under this assumption, the regressionn above identifies the average treatment effect on the treated - the average relative change in earnings in the regime with low and high taxes, for treated individuals - that is people aged 30 or below.

(d) Explain how you could provide evidence to support the crucial common trends assumption, and what statistical test you would run.

While common trends in the pre-period neither implies or is implied by common trends in the absence of treatment, it is more reasonable to believe that the trends between the groups would have been parallel after 2026 in the absence of the reform if they are also parallell in the years before the reform. To test this, we could replace the interaction term above with a full interaction between age group and year dummies, using 2025 as the base year. A joint test of whether all coefficients in this interaction for the years t=2024,2023,2022,... is a formal test of parallell trends in the pre-period.

(e) Explain how you would compute the standard errors in the difference in differences model. Why is this important?

Homoskedastic standard errors assume that the error term ε_{it} are independent and identically distributed (iid). Both parts of this assumption is unreasonable because a) it is likely that the variance of the error terms is different for people with different covariates, for instance for older and younger individuals and b) because it is unreasonable that the error terms of a particular indiviudal is uncorrelated over time. We can easily imagine that a shock to earnings (health, unemployment) in 2023 has an effect also in 2024, and ignoring this correlation would likely lead to overestimating precision /underestimating standard errors. The solution to both of these problems is to use clustered standard errors at the level of the individual, which accounts for both heteroskedasticity and correlated errors. Students could also mention the Moulton factor for this exercise to illustrate these pitfalls.

As an alternative, you instead estimate the following regression model for the year 2026 only:

$$\log(Y_i) = \alpha + \beta_0(age_i - 30) + \beta_1(age_i - 30)\mathbb{1}(age_i > 30) + \gamma\mathbb{1}(age_i \leq 30) + e_i \quad (1)$$

(f) What model is this, and what does the various components of the model do? What assumptions do you need to make for this model to estimate a causal effect? What causal effect is that?

This is a sharp regression discontinuity design model that uses age as the running variable. The β_0 term controls linearly for the relationship between age and (log) earnings before age 30, while β_1 controls for the same relationship after age 30. The parameter of interest γ estimates the jump at the threshold. For this to estimate causal effects, we need to assume continuity of potential outcomes: $\mathbb{E}(Y^1 \mid age = a)$ and $\mathbb{E}(Y^0 \mid age = a)$ is continuous in a. Intuitively, only treatment (the tax reform) can change discontinuously at the age cutoff, not potential outcomes with or without the tax reform. Under this assumption, the sharp RD parameter γ can be interpreted as the average treatment effects for 30-year olds - the average relative increase in earnings from the low tax regime.

(g) Explain what robustness checks would you perform to increase your confidence in the estimated effect and the logic behind them.

- Investigate discontinuities in other covariates at the threshold by replacing the outcome with background covariates that cannot be caused by the lower tax regime. Good candidates are variables that explain earnings, such as education. If education changes discontinuously at 30, we would worry that this is what causes any effects on labor supply rather than the tax reform.
- Investigate discontinuities in other parts of the age distribution, such as around age 25. If there are discontinuities elsewhere (where we know there is no reform), we would worry that there could be other unobservables changing also at age 30.
- Investigate robustness to making the estimates more local by a) restricting the sample to years closer to 30 or b) using weights to put more weight on observations closer to the threshold.
- Investigate robustness to higher order functional forms of the running variable.
- Perform a McCrary test of the equality of the densities at the threshold, although we would probably not expect age to be manipulable.

Students should a least mention and correctly describe the logic of 3 of these for full marks.

(h) Policymakers consider expanding the tax reform to 31-year olds. Which of the two estimates above do you find to be most informative for the suggested expansion, and why?

The fuzzy RD estimate is more local because it estimates average treatment effects for 30-year olds, whereas the DD estimate estimates average treatment effects for young people, presumably 20-30 year olds. For informing the effects of the new policy, however, this local effect is more relevant - if you think the average treatment effects for 30-year olds is more representative for 31-yar olds than the average treatment effect for all young individuals. For this reason I'd prefer the RD estimate.

3. (40 points) You have opened a webshop, and want to understand what drives traffic to your website. To do so, you log the number of visits to your site per day for a period of 10 days, and additionally record whether the day is a weekend and whether it is a day that you paid for advertising for the site on Google. You assume that the number of visits

to your website is distributed according to a poisson distribution, which has the density function

$$f(y_i;\lambda_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y!} \tag{2}$$

Where y_i is the number of visits on day i and λ_i is the conditional mean of the the number of visits on day i. You model the conditional mean as $\lambda_i = e^{\theta_0+\theta_1 ad_i}$, where ad_i is a dummy for whether you advertised on Google on day i. $y_i!$ is the factorial of y_i, so that for instance $6! = 6*5*4*3*2*1$

(a) Derive the log likelihood function from the density, and show that it may be written as $\mathscr{L}_i = \sum_i [y_i(\theta_0 + \theta_1 ad_i) - e^{\theta_0+\theta_1 ad_i} - log(y_i!)]$.

We start by finding the likelihood of observation i, then the overall likelihood is simply the product because observations (days) are independent, then take logs

$$\begin{aligned}
L_i(\theta_0,\theta_1) &= \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \\
L(\theta_0,\theta_1) &= \prod_i \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \\
\mathscr{L}(\theta_0,\theta_1) &= \log(L(\theta_0,\theta_1)) = \sum_i \log(\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}) = \sum_i [y_i \log(\lambda_i) - \lambda_i - \log(y_i!)] \\
&= \sum_i \left[y_i(\theta_0 + \theta_1 ad_i) - e^{\theta_0+\theta_1 ad_i} - \log(y_i!) \right]
\end{aligned}$$

(b) Take derivatives and interpret the first order conditions. How would you use these to find the maximum likelihood estimate of (θ_0, θ_1)?

Taking derivatives wrt θ_0 and θ_1 and setting them to zero yields

$$\begin{aligned}
\frac{\partial \mathscr{L}(\theta_0,\theta_1)}{\partial \theta_0} &\Rightarrow \sum_i y_i = \sum_i e^{\theta_0+\theta_1 ad_i} \\
\frac{\partial \mathscr{L}(\theta_0,\theta_1)}{\partial \theta_0} &\Rightarrow \sum_i y_i ad_i = \sum_i ad_i e^{\theta_0+\theta_1 ad_i}
\end{aligned}$$

The first FOC states that the MLE is found when the sum of visits to the site is equal to the sum of predicted visits to the site. The second FOC says that the sum of visits on a day with ads should be equal to the sum of predicted visits on a day with ads.
These do not have a closed form solution - despite being two equations in two unknowns, they cannot be solved to express each unknown as a function of data. Thus, we cannot find the optimum analytically and have to rely on a grid search or numerical optimization.

(c) Suppose you know that the maximum likelihood estimate is either $(\theta_0, \theta_1) = (2, 0.5)$ or $(\theta_0, \theta_1) = (1, 1)$. Using the data provided in table 1, which of them is it?

The likelihood function has three parts - $\sum_i y_i(\theta_0 + \theta_1 ad_i) = \theta_0 \sum_i y_i + \theta_1 \sum_i y_i ad_i$, $-\sum_i e^{\theta_0+\theta_1 ad_i} = -e^{\theta_0+\theta_1} \sum ad_i - e^{\theta_0} \sum (1 - ad_i)$ and $-\sum_i \log(y_i!)$. The latter has no parameters and is given in the table as -130.3. Thus we need only calculate the two first sums for the two candidate solutions, using from the table that $\sum_i y_i = 88$,$\sum_i ad_i = 4$ and $\sum_u y_{:i} ad_i = 45$

$$
\begin{aligned}
\mathscr{L}(2,0.5) &= 2\sum_i y_i + 0.5\sum_i y_i ad_i \quad -e^{2+0.5}\sum ad_i - e^2 \sum (1-ad_i) - \sum_i \log(y_i!) \\
&= 2*88 + 0.5*45 \quad -4*e^{2+0.5} - 6*e^2 - 130.3 \\
&\approx -24.86 \\
\mathscr{L}(1,1) &= 1\sum_i y_i + 1\sum_i y_i ad_i \quad -e^{1+1}\sum ad_i - e^1 \sum (1-ad_i) - \sum_i \log(y_i!) \\
&= 1*88 + 1*45 \quad -4*e^{1+1} - 6*e^1 - 130.3 \\
&\approx -43.16
\end{aligned}
$$

We can see that the $(\theta_0, \theta_1) = (2, 0.5)$ is the MLE, because the log likelihood is largest.

(d) You next add the dummy for weekend to your model and the interaction between the weekend dummy and the ad dummy, estimate the model and find that the log likelihood of the new model is -24.6. Perform a likelihood ratio test to check whether the two new terms introduced are jointly statistically significant, using significance level $\alpha = 0.05$ and critical values from table 2. What do you conclude?

The likelihood ratio test statistic is
$LR = -2(\mathscr{L}(\theta_R) - \mathscr{L}(\theta_U))$, where R and U is short for restricted and unrestricted, respectively. Using our numbers, we find that $LR = -2(24.86 - -24.6) = 0.52$. Under the null hypothesis, this test statistics is distributed as a chi-squared distribution with 2 degrees of freedom, because we added two new coefficients. The critical value for 2 degrees of freedom and significance level is 5.991 according to the table, which is much larger than our test statistic, and thus we cannot reject the hypothesis that both extra coefficients are 0.

Table 1: Website traffic data

Day (i)	**Weekend** (w_i)	**Ad** (ad_i)	**Visits** (y_i)	$\log(y!)$
1	0	1	16	30.7
2	0	0	9	12.8
3	0	0	6	6.58
4	0	0	5	4.79
5	0	1	4	3.18
6	1	1	12	20.0
7	1	0	8	10.6
8	0	0	8	10.6
9	0	0	7	8.53
10	0	1	13	22.6
Sum	2	4	88	130.3

Table 2: Critical values of chi-square distribution with df degrees of freedom

df	10%	5%	2.5%	1%	0.1%
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588