Exam 4137 – Spring 2022

IMPORTANT: Always motivate your answers. Show knowledge and understanding of the concepts taught in the course. Your answers should be as short as possible and as long as necessary. Total marks of the examination are 170. Each subquestion carries 10 marks.

---

1. Suppose you are handed a dataset with n independently identically distributed variables, Y_i. The variables only take on positive values. Assume that the draws come from a distribution with the following density:

$$f(y) = \frac{\alpha y^{\alpha-1}}{\theta^{\alpha}} \exp\left(-\left(\frac{y}{\theta}\right)^{\alpha}\right)$$

Suppose that α is known and larger than zero, but θ is unknown and therefore needs to be estimated.

(a) In words describe the logic behind maximum likelihood estimation. Under what condition is the maximum likelihood estimator consistent and efficient? Is is generally unbiased?

The maximum likelihood approach searchers for the parameters of a model whihc makes the data most likely. If the model is the correct model, Maximum Likelihood is efficient, but not unbiased. Therefore, Maximum Likelihood can only be defended on asymptotic grounds.

(b) Derive the log-likelihood function

The likelihood function is given as

$$L(\theta|\{Y_i\}) = \Pi_{i=1}^{n} \frac{\alpha Y_i^{\alpha-1}}{\theta^{\alpha}} \exp\left(-\left(\frac{Y_i}{\theta}\right)^{\alpha}\right)$$

Taking logs gives the log-likelihood:

$$\mathscr{L}_n(\theta|\{Y_i\}) = \sum_{i=1}^{n} \left(\log\alpha + (\alpha-1)\log Y_i - \alpha\log\theta - \left(\frac{Y_i}{\theta}\right)^{\alpha}\right)$$

(c) Solve for the estimator, $\hat{\theta}$. Does this estimator maximize the likelihood?

The estimator is found by solving for the first order condition.
The first derivative is given by

$$\frac{\partial \mathscr{L}_n}{\partial \theta} = \sum_{i=1}^{n} \left(-\frac{\alpha}{\theta} + \alpha Y_i^{\alpha} \theta^{-(\alpha+1)} \right) = -\frac{n\alpha}{\theta} + \alpha \theta^{-(\alpha+1)} \sum_{i=1}^{n} Y_i^{\alpha}$$

This gives the first order condition:

$$\frac{\partial \mathscr{L}_n}{\partial \theta} = 0 \Leftrightarrow \theta^{-\alpha} \sum_{i=1}^{n} Y_i^{\alpha} = n \Leftrightarrow \theta = \left(\frac{1}{n} \sum_{i=1}^{n} Y_i^{\alpha} \right)^{\frac{1}{\alpha}}$$

Note that this is positive.
The second order condition is given as:

$$\begin{aligned} \frac{\partial^2 \mathscr{L}_n}{\partial^2 \theta}|_{\hat{\theta}} &= n\alpha \hat{\theta}^{-2} - (\alpha+1)\alpha \hat{\theta}^{-(\alpha+2)} \sum_{i=1}^{n} Y_i^{\alpha} < 0 \\ &\Leftrightarrow 1 < (\alpha+1) \left(\frac{1}{n} \sum_{i=1}^{n} Y_i^{\alpha} \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} Y_i^{\alpha} \\ &\Leftrightarrow 1 < \alpha + 1 \end{aligned}$$

which is true as $\alpha > 0$.The estimate thus represent the parameter that would make it most likely that we would observe the data.

(d) You are given the following sample of Y_i in Table 1. Assume that $\alpha = 2$. What is the maximum-likelihood estimate of θ? Explain, briefly, how you would calculate the variance of the estimate. You are *not* required to calculate the estimate of the variance.

Table 1: Sample of Y_i

	1	2	3	4	5	6	7	8	9	10
Y_i	6.2	4.0	5.2	9.6	0.5	2.5	7.4	9.3	2.7	6.5

Inserting this into the estimator developed in question 1.c gives $\hat{\theta} = 6.09$. Rounding is accepted. The variance can be calculated as the inverse of the information matrix or using the outer product of the scores. Given a correct answer to question 1.c computing the estimate is fairly easy. Students should therefore be given points for concisely explaining the methods. If students have not completed quesiton 1.c they cannot compute the estimate. However, points should still be given for explaining what they would have done.

2. A government agency wants to know the effect of providing training to unemployed. They therefore run an experiment, where they allocate a group of individuals randomly into two groups. A treatment group ($G_i = 1$) receives an offer to enter a training program while a control group ($G_i = 0$) does not. The agency wishes to

measure the effect of the training program on an outcome Y_i. The government also observes whether the applicants participated in the training program ($T_i = 1$) or not ($T_i = 0$).

(a) What kind of experiment is this? What assumptions are required for estimating a causal effect of the being offered the program?

This is a randomized control trial. For a causal effect we need to assume that being assigned to the treatment group is independent of potential outcomes. If assignment to treatment is randomized, then this assumption is valid.

(b) The agency asks you to analyze the experiment. Using the variance-covariance matrix in Table 2 calculate the effect of offering the program to the unemployed. (Note: You may ignore the unit of measurement of the outcome)

The effect of offering the program is given by the slope on G_i in reduced form equation:

$$Y = \alpha_0 + \delta G_i + u_i$$

This slope can be estimated as

$$\hat{\delta} = \frac{Cov(Y_i, G_i)}{Var(G_i)} = \frac{0.07}{0.25} = 0.28$$

(c) Some of the individuals in the control group ended up participating in the training program, and some in the treatment group chose not to participate in the training program. In this light how do you interpret the effect that you estimated in question 2.b?

In case of non-compliance, the effect of group assignment is the intention to treat. Students may mention that this is the reduced form in an IV setting.

(d) What share of the population comply with the assignment to treatment and control group, i.e. take up treatment if it's offered to them?

The complier share is given by the first stage:

$$FS = \frac{Cov(T, G)}{Var(G)} = \frac{0.06}{0.25} = 0.24$$

This implies that the complier share is 24 percent.

(e) What is the effect of the training program for those who got it because the were allocated to the treatment group? What is this effect usually called? How does it differ from the average treatment effect of the treated and do they ever coincide?

Table 2: Variance-Covariance matrix

	Y_i	T_i	G_i
Y_i	1.04	0.24	0.07
T_i	0.24	0.21	0.06
G_i	0.07	0.06	0.25

This effect is called the local average treatment effect (LATE) and can be calculated with the Wald-estimator which is the reduced form divided by the first stage:

$$LATE = \frac{RF}{FS} = \frac{0.28}{0.24} = \frac{0.07}{0.06} = 1.17$$

The average treatment effect of the treated is the effect on both compliers and always takers. In the presence of always-takers the LATE and the ATT will not coincide. However, if applicants in the control group can be perfectly prevented from obtaining training, then the two effects coincide.

3. During the Covid-19 pandemic policymakers have implemented a wide range of non-pharmaceutical interventions to reduce the number of infections. An important question is whether such policies effectively reduce infections.

   (a) Many countries decided to implement a nationwide lockdown in March 2020. However, Swedish officials chose not to implement a nationwide lockdowns. You have access to weekly infection rates for multiple countries. Your colleague suggest comparing Sweden to countries without lockdowns to estimate the effect of the lockdowns by running the following regression on data from the first half of 2020:

   $$Y_{it} = \beta_0 + \beta_1 Sweden_i + u_{it},$$

   where Y_{it} denotes infection cases in country i at week t and $Sweden_i$ is a dummy which is 1 if country i is Sweden. What would β_1 capture? Does this capture the effect of lockdowns? Discuss why/why not.

   $\beta_1$would capture the average difference between infection rates in Sweden and the other countries in the sample in the first half of 2020. This includes the months before the lockdown where infections were relatively low in all countries. If Sweden differs in other respects than the lockdown policy this would be reflected in β_1. Obvious candidates for such factors are geography, demography, climate and political system.

   (b) Another colleague suggest that you instead run the following regression:

   $$Y_{it} = \lambda_0 + \lambda_1 Sweden_i + \lambda_2 AftMarch_t + \lambda_3 AftMarch_t \times Sweden_i + u_{it},$$

   where $Sweden_i$ is a dummy which is 1 if country i is Sweden and 0 otherwise. $AftMarch_t$ is a dummy which is 1 when the observation is recorded in March

or later. What kind of research design is this? Interpret the coefficients in this regression. Under which conditions can this model be used to measure the effect of lockdowns? Discuss.

λ_1 captures the average difference in infection rates between Sweden and the other countries before March. λ_2 captures the average difference in infection rates before and after March excluding Sweden. λ_3 captures the excess difference for Sweden. If we are willing to assume that the only reason why the trend in infections in Sweden would differ from other countries is because of the lack of lockdown measures, then λ_3 would capture the effect of lockdowns (with a sign reversal). This is a difference-in-difference design. However, if the trend in Sweden differed for other reasons, then the design is invalid. If for instance the choice of lockdowns were a function of trends, this would violate the common trend assumption. Points should be given for explicitely discussing the context of the common trend assumption in this setting.

(c) One of the countries in your dataset is Denmark. Malmø, the third largest city in Sweden, and the Danish capital Copenhagen, are part of the same metropolitan region with many people commuting across the border. Denmark chose to lock down mid-March. Is this a problem for the research design in question 3.b? (Hint: Discuss SUTVA)

Yes. The Danish lockdown will affect commuters and part of Sweden will therefore be "treated" with a lockdown. This case of spill-overs breaks the SUTVA-assumption.

(d) In a paper in the Econometrics Journal, Cho, S. W. (2020) seeks to estimate the effect of lockdowns.[1] Using weekly data on infection rates for Sweden along with Finland, Greece, Norway, Denmark, and Estonia. All countries beside Sweden implemented lockdowns at some point in the sample period. Cho estimates the following model:

$$Y_{it} = \eta_i + \nu_t + \theta L_{it} + \varepsilon_{it},$$

where Y_{it} denotes infection cases per million inhabitants in country i at time t, η_i and ν_t are country and time fixed effects respectively, and L_{it} is a binary treatment dummy which equals 1 if country i was locked down at time t. The treatment dummy will equal 1 for at least one period for all countries except Sweden. How does this model differ from the model in question 4.b? What does an estimate of θ measure? Under what conditions can θ be given a causal interpretation.

[1] Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. The Econometrics Journal, 23(3), 323-344.

Table 3: Results from Cho (2020)

	Infection rate (Y_{it})
Lockdown (L_{it})	-751.1
S.E.	(90.75)
N	600
R^2	0.773

The model is a generalization of the difference in difference model in question 4.b. The country fixed effect allows for a different intersect for each country. The time-fixed effects allow for a fully flexible time trend share across observations. Given a common trend assumption, θ captures the causal effect of lockdown measures on infection rates. The students may discuss questions of timing, such as measurement, incubation period etc. This is not required, and the text in the question is a bit vague. Considerations of such question should weigh positively in the assessment of the response.

(e) How can the equation in 3.d be estimated? Why is there no constant?

The model can be estimated with Least Square Dummy Variables method (LSDV) or the within estimator/fixed.effects estimator or by the first-difference estimator. Either way you estimate it the constant, which does not vary over individuals or time will drop out.

(f) The results of this estimation is presented in Table 3. Interpret the result. Is the R^2 in line with expectations and is it important? Is the parameter on L_{it} significant?

Introducing a lockdown measure is associated with a fall in the infection rate of -751. The effect is highly significant with a $t = \frac{-751.1}{90.75}$. We include both time and country fixed effects and therefore exclude a lot of variation in infection rates. Therefore it is unsurprising that the explained variance is quite high. As we are interested in a causal effect we do not worry too much about a high R^2.

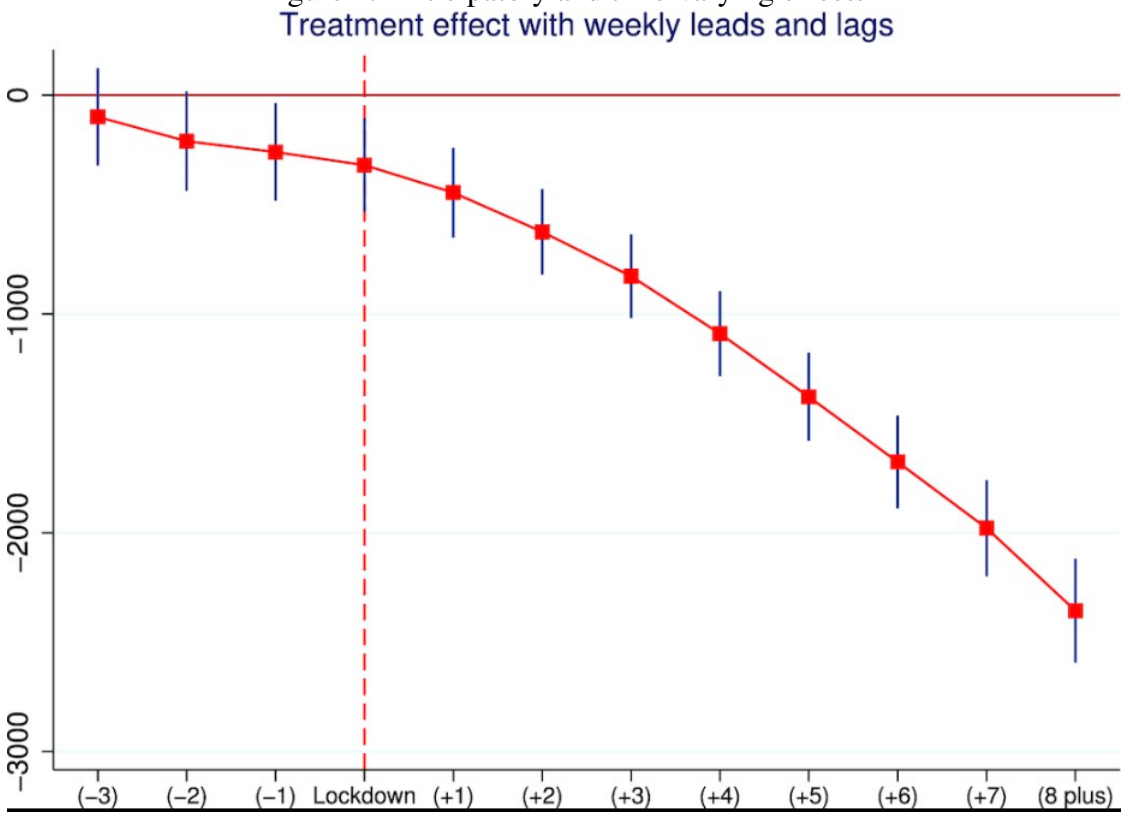
(g) Cho writes "*Although most [...] applications customarily cluster standard errors at the aggregate level, there are too few units of observation to follow clustering methods. As such, I report heteroscedasticity robust standard errors.*". In light of this quote, reflect on your answer to the previous question.

Difference-in-Difference designs are generally plagued by correlations among observations over time and within groups, here countries. This means that it is important to consider clustering of standard errors. The implementation by Cho likely underestimates the standard errors which makes us more likely to reject a true null (Type I error). Students may mention the Moulton factor and the fact that we follow the same observations over time.

(h) Cho presents the figure displayed in Figure 1. What does this figure show?

How would you assess the research design in light of this figure? Discuss.

Figure 1: Anticipatory and time-varying effects



Notes: Horizontal axis shows weekly leads and lags from the day of lockdown. Vertical bands are 95 percent confidence bands. Source: Econom J, Volume 23, Issue 3, September 2020, Pages 323–344, https://doi.org/10.1093/ectj/utaa025

The figure shows clear indications of a downward trend before the implementation of the lockdown and no sharp fall at the moment of the lockdown. This does not provide convincing evidence that the difference-in-difference design is appropriate. However, the students may again discuss timing of measurement and contagion. Also the students may discuss anticipatory effects and relate it to Ashenfelter's Dip.