Reconstructie Econometric Methods and Models Januari 2022

10 questions, almost all based on sample exam and previous exam questions. You got 3 hours.

1. **10 points** Assume that the true data generating process (the true model) is: $y = 20 - 30(x + r) + 4w + \epsilon$. Suppose that we also know that $\mathrm{E}(\epsilon|x) = \mathrm{E}(\epsilon|w) = \mathrm{E}(\epsilon|r) = 0$, $Cov(x, w) = 20$, $Cov(x, r) = Cov(w, r) = 0$, $Var(x) = 10$, $Var(w) = 8$ and $Var(r) = 6$. What would be the estimate of the coefficient for the variable x if we regress y on a constant and the variable x via OLS and we have a very large sample? Carefully motivate your answer.

   Very similar question, only cov(*x*, *r*) = cov(*x*, *w*) = 5. Cov(*w*, *r*) = 10
2. The proof of consistency of Maximum Likelihood estimation has two parts. Could you provide the intuition behind each of those parts for the exponential model? Carefully motivate your answer.
3. Assume the equation Yi = α+xiβ + εi. With E(ε)=0 and E(ε|x)=/= 0. There is also a variable zi where E(εzi)=0. What extra assumptions do you need to make to estimate this model in GMM? What are the moments in GMM? Give the steps (give extra assumptions if needed, otherwise don't). Carefully motivate your answer
4. We have $y_i = 1$ if $y^*_i > 0$ and $y_i = 0$ else, with $y^*_i = \alpha + x_i\beta + \varepsilon_i$. ε is normally distributed with mean 0 and variance 4. Can we get a consistent estimate for the parameter for *x* if we use a Probit model with $\varepsilon \sim N(0, 1)$? Can we get a consistent estimate for the value of α relative to β?
5. Assume that there are two individuals (A,B) and we observe them in two periods (1, 2). Provide the full expression of the numerator of the estimated *clustered standard errors in panel data* for the cases in which a) we cluster by time period, b) we two-way cluster by individual and time. Recall that the formula for the estimated *variance in panel data* is given by the following equation. Note that this is for the *variance*, not the *standard error*.

$$\hat{VAR}(\hat{\beta_{ols}} - \beta) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}(X_{it} - \bar{X}_i - \bar{X}_t)(X_{js} - \bar{X}_j - \bar{X}_s)e_{it}e_{js}\mathbb{1}[\text{i,t,j,s in same cluster}]}{(\sum_1^T \sum_1^N (X_{it} - \bar{X}_i - \bar{X}_t)^2)^2}$$

6. Assume that we randomize the eligibility to participate in a training program for individuals. In order to evaluate the program, we are given the following information on the average wages for a group of individuals that participated in the program (Treated Individuals) and a group of individuals that did not participate (Control Individuals).

	Before	*After*
Treated individuals	9	13
Control individuals	12	12

Suppose we estimate the following model: $Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$. What is your estimate for β? Explain under which assumptions this estimate is consistent. Can we estimate the α_i's?

7. Explain under what conditions panel data (in FE) can increase bias due to measurement error. Carefully motivate your answer.
8. What is the difference between difference-in-difference estimation and time and individual effects estimation?
9. Consider the following model: $y_t = \varrho y_{t-1} + \epsilon_t$. Explain under which assumptions we can consistently estimate rho. What steps would we need to take? What if this model is a random walk? Can we test whether this is a random walk?
10. Describe in your own words what Heckman's selection model is and what kind of problems does it solve? And what are its main limitations?