**5304 Econometrics Autumn 2023**[1]

**Problem set 6**
**Old exam question**

---

This problem set consists of an old exam question, and thus does not include any data work. The question is from the 2017 take-home exam, and deals with issues related to panel data. Please submit typed solutions as per usual.

[1]TA: Petter Berg, petter.berg@phdstudent.hhs.se

## Question B: The quality of hospital care for elderly patients (40 points)

A team of researchers want to estimate the quality of different hospitals in a large European city in treating patients over 65 years. They decide to focus on treatment after emergency care and choose patient mortality one year later as their primary outcome measure for comparing hospitals. They receive administrative data on the universe of patients who went to each of 10 hospitals in the city over the past 10 years, including their past medical histories, their medical diagnosis after approaching the hospital for emergency care, and various demographic data such as age (in months) and sex. They begin with estimating the following regression:

$$M_{ih} = \alpha + \sum_{j=2}^{10} \mu_j . I[h = j] + \gamma . Z_{ih} + \epsilon_{ih} \quad (1)$$

where $M_{ih}$ is a dummy variable for whether the patient $i$, who went to hospital $h$, was alive one year after s/he first went to the hospital for emergency treatment. $Z_{ih}$ includes controls for patient gender, a linear control for age, whether the patient smoked and whether s/he was obese. The coefficients of interest are given by $\mu_j$ where, e.g., $\mu_2$ is the coefficient on the dummy variable of having gone to hospital 2. If a patient had multiple emergency visits/hospitalization, the dataset is aggregated to have only one observation for the patient starting from their first visit. For simplicity, assume that each patient has only approached one hospital for emergency care ever.

1. Computing hospital quality:
   (a) Under what assumption would $\mu_j$ be interpretable as the difference in quality of care between hospital $j$ and hospital 1 (the omitted hospital)? Assume at least some patients could choose which hospital to go to within the city and previous studies show that patients often prefer to go to nearer hospitals in emergencies. In this scenario, is the assumption needed above likely to be satisfied or not? Explain your reasoning. [10%]
   (b) The researchers discover that some hospitals have specialist care for particular medical conditions (e.g. cancer) and therefore are more likely to be referred to patients with those diseases/conditions i.e. the composition of cases by medical condition differs across hospitals. Under what circumstances would that be a problem for interpreting $\mu_j$ as the causal difference in hospital quality? Could you use (some of) the other data that the researchers have access to to reduce this concern? If yes, how would you propose to do that? [10%]
   (c) A medical colleague points out that a linear control for age is unlikely to be suitable because the risk of dying increases substantially faster with age the older you get. Assuming that the researchers have a large sample of individuals with coverage from 65-95 years of age, what would be your recommendation for how to address this concern and why? [10%]
   (d) The same medical colleague points out that one important cause of mortality risk is whether or not medical personnel checked early on for signs of a heart attack. The researchers observe two dummy variables in their dataset on this topic: (a) whether this check was administered by ambulance staff ($Checkambu = 1$) and (b) whether this check was administered by the medical staff in the emergency room ($CheckER = 1$). Should they include these variables in $Z_{ih}$? If not, how should they use this? Assume that, in this city, ambulance services are run independently and are not managed by hospitals directly. [10%]
2. The researchers are concerned about the differential case load between hospitals (such as in part 1b above) and attempt to address it using an instrumental variable strategy.
   (a) State the conditions needed for a variable to be an appropriate IV in this case. How many such instruments do the researchers need to find? [Note: there are multiple hospital dummy variables in Eq 1] [10%]
   (b) The researchers compute, for each neighbourhood in the city and year, the most common hospital that patients from that neighbourhood went to in the five years preceding. They want to use this variation to predict where a given patient is likely to go. How would you use this variation to create the

required instrumental variable/s? Provide the first-stage and structural/second-stage equations. Is this IV strategy likely to be valid? Explain why or why not. [15%]

(c) The researchers note that there are multiple ambulance providers in the city, which may offer a potential IV strategy. For simplicity, assume that all patients going to emergency care arrived by ambulance.

i. City rules indicate that the ambulance that responds to an emergency call is assigned quasi-randomly. Specifically, when a patient calls emergency services, they dispatch the nearest available ambulance (without considering which provider operates the ambulance). Which ambulance provider is available is assumed to not be related to patient characteristics. The researchers wish to verify that which ambulance service is available is not related to patient characteristics – how would you recommend that they do this? [5%]

ii. The researchers note that while all ambulance providers serve multiple hospitals, providers have different probabilities of taking patients to any given hospital.[1] They thus intend to use the quasi-random variation in which ambulance provider picked a patient up to deal with the potential selection bias in hospital choice. How would you recommend that they do this? Provide both the first stage and structural/second-stage equations. [10%]

(d) The researchers note that, outside of emergency settings, patients have to go to the hospital that serves their neighbourhood and each hospital has a well-defined service area. The researchers note that area boundaries are typically demarcated by a street (so, e.g., one side of the street is in the service area of Hospital 3 and the other side is in the service area of Hospital 4). They have the intuition that nothing else varies across two sides of street than this arbitrary line. How would you test this intuition and, assuming their intuition is correct, could you use this variation to create appropriate IV(s) for hospital choice? [10%]

(e) The researchers arrive at IV estimates using the strategy in both 1c and 1d and find that these differ from OLS estimates and that they differ from each other. Assuming that both IV strategies in 1c and 1d were valid and had a strong first stage, how would you interpret these differences? [10%]

[1] For example, based on historical data, Ambulance company A has a 30% probability of taking patients to Hospital 1, 20% to Hospital 2, 15% to Hospital 3 and 5% to each of the other hospitals. Similarly, each of the other companies has a different probability of taking patients to a given hospital.