1 IV questions

1. **Instrumental variables I.** In Acconcia et al. (2014), "Mafia and Public Spending", American Economic Review 104(7), the authors use Italian data on 95 provinces (indexed by i) to estimate the regression coefficient β in the model

$$Y_i = \beta G_i + \gamma X_i + v_i$$

where Y_i is the growth of per capita value added in a province i; G_i is per capita spending on infrastructure in the same province; and X_i denotes further controls measuring the number of people reported to the judicial authority for (i) Mafia-type association, (ii) extortion; (iii) Mafia-related murders; (iv) corruption; and (v) the number of corruption crimes reported to the judicial authority.

   (a) Explain why $G_{i,t}$ is endogenous. *Provinces that are not doing well are more likely to be allocated aid from the central government.*

   For their instrumental variable strategy, the authors exploit a new anti-mafia/corruption policy. The new policy is such that, upon discovering mafia involvement, the city council is dismissed, and financial flows into local public works and investment projects (where the mafia involvement typically occurs) is cut. This allows the province to have a fresh, mafia-free start. The instrument Z_i is "number of city councils dismissed in province i".

   (b) Do you think that Z_i is a relevant instrument? *Yes, since discovery of mafia involvement is followed by cuts in spending, G is expected to be affected. This can be checked.*

   (c) Convince me that Z_i is a valid instrument. *Note that X already contains a lot of information about the presence and size of mafia, so the instrument is only about the discovery of them. Many other correct answers are possible: you need to tell a story about the relationship between the instrument and the error term v_i.*

2. **Instrumental variables II.** Evans and Schwab (1995) studied the effects of attending a Catholic high school on the probability of attending college. For concreteness, let *college* be a binary variable equal to unity if a student attends college, and zero otherwise. Let $CathHS$ be a binary variable equal to one if the student attends a Catholic high school. The multiple linear regression model is:

$$college = \beta_0 + \beta_1 CathHS + other\ factors + u$$

where the other factors include gender, race, family income, and parental education.

   (a) Why might $CathHS$ be correlated with u? **Catholic schools are not so different from private schools. Parents who make the effort of sending their kids to private schools will also be more likely to support their child to go to college.**

   (b) Standardized test scores from the students' second year in high school are available. What can be done with this variable to improve the estimate of the causal effect of attending a Catholic high school? **Added as a control variable!**

(c) Let $CathRel$ be a binary variable equal to one if the student is Catholic. Discuss the two requirements needed for this to be a valid instrumental variable for $CathHS$ in the preceding equation. **It would have to be correlated with CathHS, and it should not be correlated with u (it should affect college only through CathHS and the other regressors.)**

(d) Which one of these two requirements can be tested? **That it is correlated with HS**

(e) Being Catholic has a significant effect on attending a Catholic high school. Do you think $CathRel$ is a convincing instrument for $CathHS$? **That depends on what is captured by u. Say that a spiritual/religious background makes students more introspective and curious, or that this background trains your concentration through prayer. This may make it more likely that a student is interested in college, or that they will do well in high school. This is not captured by HS, and may violated the exogeneity of the instrument.**

2 Cheat sheet

1. Critical values:

   (a) 5%: 1.96

   (b) 1%: 2.58

2. Adjusted R-squared:

$$\begin{aligned} \bar{R}^2 &= 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} \\ &= 1 - \frac{n-1}{n-k-1}\left(1 - R^2\right). \end{aligned}$$

3. Two stage least squares estimator with a single regressor and a single instrument

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}}.$$