Ph.D ECON 510. August 2019
ECONOMETRICS CANDIDACY EXAM

Andres Aradillas-Lopez

Consider the following structural equation

$$Y_{1i} = X_i'\beta_0 + \delta_0 Y_{2i} + \varepsilon_i. \tag{1}$$

Let $\theta_0 \equiv (\beta_0', \delta_0)'$ denote the structural parameters. Y_{2i} is a scalar random variable and X_i is a random vector. Assume X_i is exogenous: $E[X_i \cdot \varepsilon_i] = 0$, but $E[Y_{2i} \cdot \varepsilon_i] \neq 0$ (i.e, Y_{2i} may be endogenous). Suppose we model the endogeneity of Y_{2i} through the following reduced-form equation:

$$Y_{2i} = Z_i'\gamma_0 + \eta_i, \tag{2}$$

Suppose Z_i is exogenous in the following sense: $E[Z_i \cdot \eta_i] = 0$ and $E[Z_i \cdot \varepsilon_i] = 0$. Also, suppose $E[X_i \cdot \eta_i] = 0$. Assume throughout that Z_i contains at least one variable excluded from X_i. We allow for $E[\eta_i \cdot \varepsilon_i] \neq 0$, and therefore η_i is the source of endogeneity for Y_{2i} in (1). Plugging in (2) in (1), we have

$$\begin{aligned} Y_{1i} &= X_i'\beta_0 + \delta_0 \cdot [Z_i'\gamma_0 + \eta_i] + \varepsilon_i \\ &\equiv X_i'\beta_0 + \delta_0 \cdot Z_i'\gamma_0 + \nu_i, \end{aligned} \tag{3}$$

where $\nu_i \equiv \delta_0 \cdot \eta_i + \varepsilon_i$. Suppose we have an iid sample $(Y_{1i}, Y_{2i}, X_i, Z_i)_{i=1}^n$.

(1) **[60 points]** Suppose we estimate θ_0 through a two-step procedure:

> **Step 1:** We estimate γ_0 in (2) by OLS. Label the estimator as $\widehat{\gamma}$.
>
> **Step 2:** We estimate θ_0 by running an OLS regression of (3) replacing γ_0 with its OLS estimator $\widehat{\gamma}$ from the first step.

Label the resulting estimator from this two-step procedure as $\widehat{\theta}^{2SLS}$ (two-stage least squares). Describe the asymptotic properties of $\widehat{\theta}^{2SLS}$ by characterizing its influence function.

(2) **[40 points]** Suppose we ignore the endogeneity of Y_{2i} and we simply go ahead and estimate θ_0 by running an OLS regression of the structural equation (1). Label the estimator as $\widehat{\theta}^{OLS}$.

(i) **[20 points]** Using (2), describe the probability-limit of $\widehat{\theta}^{OLS}$, and label it θ^*.

(ii) **[20 points]** Once again, using (2), describe the linear representation of $\widehat{\theta}^{OLS}$.

Please note that your answers to (i) and (ii) require you to use the reduced-form equation (2).

Patrik Guggenberger

Comprehensive Exam 2019

Please provide complete explanations for your answers and write legibly. 60 points total.

1. **[9 points]** Let the joint density function of X and Y be

$$f(x,y) = \begin{cases} \frac{2}{(1+x+y)^3} & \text{for } 0 < x < \infty \ \&\ 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}.$$

(a) Compute the marginal density of X.

(b) Compute the conditional density of Y given $X = x$ for $x > 0$.

(c) Compute $E(Y|X = x)$ for $x > 0$.

2. **[9 points]** Suppose X is a random p-vector on a probability space $(\Omega, \mathcal{F}, P)$ and $g : R^p \rightarrow R$ is a measurable function such that $\int |g(x)|dP < \infty$. Recall that $Eg(X) = \int g(X)dP$. Show that $Eg(X) = \int g dP_X$, where P_X is the distribution of X.

3. **[12 points]** Let $X_1, ..., X_n$ be i.i.d. random variables with mean μ and variance σ^2 and define $C = \sigma/\mu$. Let

$$\begin{aligned} \widehat{C}_n &= \widehat{\sigma}_n/\overline{X}_n, \text{ where} \\ \overline{X}_n &= n^{-1}\sum_{i=1}^{n} X_i \text{ and} \\ \widehat{\sigma}_n^2 &= n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2. \end{aligned}$$

(a) Establish the consistency of $\widehat{C}_n$ as an estimator of C. (You can use the law of large numbers, rules concerning convergence in probability, and Slutsky's Theorem.) State any assumptions you use to obtain the result.

(b) Find the asymptotic distribution of

$$n^{1/2}\begin{pmatrix} \overline{X}_n - \mu \\ \widehat{\sigma}_n^2 - \sigma^2 \end{pmatrix}.$$

Again, state any assumptions you use to obtain the result.

4. **[10 points]** Let $X_1, ..., X_n$ be i.i.d. mean zero random variables with distribution that is symmetric about zero and with variance that is not necessarily finite. Prove the law of large numbers for these random variables by a truncation argument. Such an argument uses the fact that we can write

$$\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i 1(|X_i| \leq c) + n^{-1}\sum_{i=1}^{n} X_i 1(|X_i| > c)$$

for any constant $c > 0$.

5. **[12 points]** Suppose $X \sim bin(n, p)$ with probability mass function $f(x, p) = \binom{n}{x} p^x(1-p)^{n-x}$ for $x = 0, ..., n$ and $p \in [0, 1]$. One observes a single random variable X.

   (a) What is the maximum likelihood estimator of p?

   (b) What is the asymptotic distribution of the ML estimator of p as $n \to \infty$?

   (c) State the Neyman-Pearson Lemma.

   (d) Find the level α uniformly most powerful test of test of $H_0 : p = p_0$ versus $H_1 : p = p_1$ for $p_1 > p_0$. Simplify the form of the test as much as possible. (Although X is discrete, for the purposes of this question and the question in part (d), you do not need to worry about that when specifying the critical value. That is, specify the critical value as if the distribution was continuous.)

   (e) Find the level α UMP test of $H_0 : p \leq p_0$ versus $H_1 : p > p_0$.

6. **[8 points]** Let Y_n be a sequence of random variables s.t. $\sqrt{n}(Y_n - \theta) \to_d Z \sim N(0, \sigma^2)$ as $n \to \infty$ for some scalars θ and $\sigma^2 > 0$. Let $g \in C^2(U)$, where U is an open neighborhood of θ, a scalar valued function. Assume $g'(\theta) = 0$. Work out the limiting distribution of $g(Y_n)$ (First find the correct normalization and centering). Clearly state any additional assumptions you need.