# ASSIGNMENT 2

## ETHAN LIGON

You are strongly encouraged to work as a team, and to turn in a single assignment for grading. The principal deliverable you turn in should be a link to a `github` repository. Please complete work on **four** of the following six sections, your choice (but all of this material is fair game for the final).

### 1. EXERCISES (IDENTIFYING ASSUMPTIONS FOR REGRESSION)

(1) Evaluate the truth of following statement: "In the linear regression $y = X\beta + u$ the usual identifying assumption $\mathbb{E}(u|X) = 0$ (call this an assumption of "mean independence") implies $\mathbb{E}(h(X) \cdot u) = 0$ for any function $h$ satisfying some regularity conditions related to measurability."

(2) Suppose $y$, $x$ and $u$ are scalar random variables, with $y$ and $x$ observed but $u$ unobserved. Consider the function $h(x) = x^3$; under standard assumptions this satisfies our concerns about measurability, so $\mathbb{E}(u|x) = 0$ implies $\mathbb{E}(ux^3) = 0$. Use this last condition to motivate a simple least squares estimator of the regression equation $y = \alpha + \beta x + u$. How does this differ from the usual OLS estimator? Why might one prefer one to the other, and under what conditions?

(3) Sometimes we will encounter estimators (e.g., Maximum likelihood) that adopt an assumption of *independence*, rather than mean independence. In the current setting this might be expressed as something like $\Pr(x < x \cap u < u) = F(x)G(u)$ for some cumulative distribution functions $F$ and $G$. Show that independence implies mean independence, but not the converse.

(4) Related to the previous: Show that while $u$ mean independent of $x$ implies $\mathbb{E}(uh(x)) = \mathbb{E}(u) = 0$, independence also implies $\mathbb{E}(g(u)x) = \mathbb{E}x\mathbb{E}g(u)$.

(5) Suppose that $y = f(X) + u$ for some unknown but continuous function $f$. Suppose we want to use observed data on $X$ to predict outcomes $y$, and seek a predictor $\hat{y}(X)$ which is "best" in the sense that the (expected) mean squared prediction error

$\mathbb{E}[(y - \hat{y}(X))^2 | X)$ is minimized. What can we say about $\hat{y}$ and its relation to the conditional expectation $\mathbb{E}(y|X)$? Its relation to $u$?

(6) Let $y = X\beta + u$, and let $D$ be a binary random variable. with $\mathbb{E}(u|D) = 0$ and $\mathbb{E}(X|D) \neq \mathbb{E}(X)$. Establish that $D$ is a valid instrument, and work out a particularly simple expression for the IV estimator in this case. Discuss.

(7) Write out the two causal diagrams which justify, respectively, the least squares estimator and the IV estimator. What would it mean for one model to be correct, but not the other? How could you test this?

## 2. WRIGHT (1928)

Consider the canonical demand and supply model in which quantity supplied is a function of price and a set of "supply shifters"; quantity demanded is a function of price and set of "demand shifters"; and market clearing implies that at some price quantity demanded is equal to quantity supplied.

A linear version of this model is fully specified and solved in this Jupyter Notebook.

Consider the following questions:

(1) (Control) What is the expected demand if we *set* the price $p = p_0$?

(2) (Condition) What is the expected demand if we *observe* $p = p_0$?

(3) (Counterfactual) If prices and quantities are observed to be $(p_0, q_0)$, what **would** demand be if we **were** to *change* the price to $p_1$, *ceteris paribus*?

Answers could be mathematical expressions, or code that answers the question for the model given in the Jupyter notebook.

## 3. "PLAUSIBLY EXOGENOUS"

The Wright (1934) model we've described takes the form

$$(1) \qquad\qquad\qquad y = X\beta + u$$

with a right-hand-side variable (price) that depends on the disturbance $u$.

Were we to estimate the *regression* equation $y = Xb + e$ using least squares, we would obtain $b = X^+ y = \beta + X^+ u$; we cannot *identify* $b$ with $\beta$ because of the unknown and unobservable term $X^+ u$.

(1) We previously found that with some instrumental variables $Z$ satisfying the moment condition $\mathrm{E}(u^\top Z) = 0$, it becomes possible to identify $\beta$. Explain in detail how this works, and exactly what assumptions are required.

Conley et al. (2012) describe methods for dealing with cases in which the moment condition $\mathrm{E}(u^\top Z) = 0$ is violated, so that in a regression

$$(2) \qquad\qquad\qquad y = X\beta + Z\gamma + u$$

estimates of the parameter vector $\gamma$ may be non-zero, but where $\gamma$ is nevertheless "small".

(1) Modify the framework we developed for exploring the Wright (1934) model so that the data generating process `linear_dgp` allows for $\gamma \neq 0$, and explore how estimator $b(\gamma)$ varies with $\gamma$.
(2) Calculate a region $A$ over which one might *fail to reject* (i.e., "Accept") the null hypothesis that $b(\gamma) = b(0)$ at a conventional level of significance. Discuss.
(3) Further modify the framework so that the covariance of $Z$ and $X$ is equal to a parameter $\sigma_{XZ}$. Calculate the set $B$ of pairs $(\gamma, \sigma_{XZ})$ such that one would fail to reject the same null hypothesis at the same level of significance. Discuss.

## 4. Weak Instruments

This problem explores the problem of weak instruments. The basic setup should be familiar, with

$$(3) \qquad\qquad \begin{aligned} y &= \beta x + u \\ x &= Z\pi + v \end{aligned}$$

. Note that we've assumed that $x$ is a scalar random variable, and that $Z$ is an $\ell$-vector. (In general we might have $k$ endogenous $x$ variables, so long as we have $\ell > k$.)

(1) Construct a data-generating process `dgp` which takes as arguments $(n, \beta, \pi)$ and returns a triple $(y, x, Z)$ of $n$ observations.
(2) Use the `dgp` function you've constructed to explore IV (2SLS) estimates of $\beta$ as a function of $\pi$ when $\ell = 1$ using a Monte Carlo approach, assuming homoskedastic errors.
  (a) Write a function `two_sls` which takes as arguments $(y, x, Z)$ and returns two-stage least squares estimates of $\beta$ and the standard error of the estimate.
  (b) Taking $\beta = \pi = 1$, use repeated draws from `dgp` to check the bias, and precision of the `two_sls` estimator, as well as the size and power of a $t$-test of the hypothesis that $\beta = 0$. Discuss. Does a 95% confidence interval (based on your

2SLS estimator) correctly cover 95% of your Monte Carlo draws?

(c) Taking $\beta = 1$, but allowing $\pi \in [0, 1]$ again evaluate the bias and precision of the estimator, and the size and power of a $t$-test. The $Z$ instrument is "weak" when $\pi$ is "close" to zero. Comment on how a weak instrument affects two-stage least squares estimators.

(3) Now consider another "weak" instruments problem. Consider the sequence $\{1, 1/2, 1/4, 1/8, \dots\}$. Let $\ell = 1, 2, 3, \dots$, and for a particular value of $\ell$ let the vector of parameters $\pi_\ell$ consist of the first $\ell$ elements of the sequence. Thus, your `dgp` should now return $Z$ we can treat as an $n \times \ell$ matrix, with successive columns of $Z$ increasingly "weak" instruments.

(a) Taking $\beta = 1$, but allow $\ell$ to increase ($\ell = 1, 2, \dots$). Note that for $\ell > 1$ this is now an "overidentified" estimator. Describe the bias and precision of the estimator, and the size and power of a $t$-test. Compare with the case of $\ell = 1$ and $\pi = 1$.

(b) What can you say about the optimal number of instruments (choice of $\ell$) in this case?

## 5. A SIMPLE APPROACH TO INFERENCE WITH WEAK INSTRUMENTS

Chernozhukov and Hansen (2008) propose a very simple way to handle inference in a linear IV model, even in the case in which instruments are many and/or weak. This problem explores the problem of weak instruments, and their method of inference. The basic setup should be identical to the above, with

$$(4) \qquad \begin{aligned} y &= \beta x + u \\ x &= Z\pi + v \end{aligned}$$

. In this problem you will use the same `dgp` as in the previous problem.

The idea of Chernozhukov and Hansen is simple: If we can specify a regression in which all the endogenous variables are on the left-hand side, then OLS is consistent. So, they subtract $\beta_0 x$ from both sides of the estimating equation (for some choice of $\beta_0$), and then use the expression for $x$ to substitute using $Z$, or

$$\begin{aligned} y - \beta_0 x &= x(\beta - \beta_0) + u \\ y - \beta_0 x &= (Z\pi + v)(\beta - \beta_0) + u \\ y - \beta_0 x &= Z\gamma + w. \end{aligned}$$

The key is that if $\beta_0 = \beta$, then we will have $\gamma = 0$. So the idea is to try to find $\beta_0$ such that OLS estimates of $\gamma$ in $y - \beta_0 x = Z\gamma + w$ are close to zero.

(1) Again suppose that the true $\beta = 1$. Write a function which takes as arguments $(y, x, Z, \beta_0)$ and which returns the $p$-value associated with the hypothesis that every element of $\hat{\gamma}$ is zero (an $F$-test would be appropriate). Note that this same $p$-value characterizes the hypothesis test that $\beta = \beta_0$.

(2) Using your function and taking $\pi = 1$, estimate $\beta$ by finding the value of $\beta_0$ which delivers maximal $p$-values. Describe the bias and precision of this estimator.

(3) Use the fact we've described about $p$-values above to construct 95% confidence intervals for your estimator of $\beta$. Consider the coverage of this 95% confidence interval, as in the previous question. How does this compare with the 2SLS case?

(4) What happens to the coverage of your test as $\pi$ goes from 1 toward zero? How does this compare with the 2SLS case?

(5) Using the same construction of "many instruments" as in the previous question, how does the coverage of your test change as $\ell$ grows large? Again, compare with 2SLS.


## 6. ANGRIST-KRUEGER (1991) REPLICATION

You'll find data from a famous paper by Angrist and Krueger (1991) in the ARE212_Materials repository, along with a pdf of the paper. The paper uses information on "quarter of birth" as an instrument for (endogenous) education to measure returns to education.

The first specification in the paper is given in their equations (1) and (2).

(1) What is the (implicit) identifying assumption? Comment on its plausibility.

(2) Using their data, estimate (2), replicating the figures in their Table 5, using the conventional two-stage least squares IV estimator (what they call TSLS).

(3) Repeat (2), but for the specification reported in their Table 7 (which has many more instruments). Summarize what the above exercises tell us about returns to education.

(4) Adapt your implementation of the Chernozhukov and Hansen estimator to estimate the key parameter $\rho$, first for the Table 5 specification, then the Table 7 specification. How does your point estimate compare?

(5) Same as (4), but construct 95% confidence intervals using both 2SLS and your new estimator. How do these compare? Which estimator do you prefer, and why?

Two useful pandas tricks to know for this exercise: `pandas.read_stata`, `pandas.get_dummies`.

## REFERENCES

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Chernozhukov, V., & Hansen, C. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, *100*(1), 68–71. https://faculty.chicagobooth.edu/christian. hansen/research/ch_weakiv_mar07.pdf

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly exogenous. *Review of Economics and Statistics*, *94*(1), 260–272.

Wright, P. G. (1928). *Tariff on animal and vegetable oils*. Macmillan Company, New York.