ARE 212

Problem Set 2, Second Half

Scott Kjorlien, Thor Larson, Sarah O'Brien, Bora Ozaltun, Garrison Schlauch

1 1. Exercises

Part (1)

True.

Proof:

$$\begin{aligned}\mathbb{E}[h(X)u] &= \mathbb{E}_X\{\mathbb{E}[h(X)u|X]\} \quad \text{by the law of iterated expectations}\\ &= \mathbb{E}_X\{h(X)\,\mathbb{E}[u|X]\} \quad \text{since we're conditioning on } X\\ &= 0 \quad \text{since } \mathbb{E}[u|X] = 0 \text{ by assumption}\end{aligned}$$

Part (2)

Consider the general least squares model:

$$y_i = X_i'\gamma + u_i \quad E[u_i|X_i] = 0$$

where X_i and γ is a $k \times 1$ vector. Consider any nicely behaved transformation of our data: $g(X_i)$ (also a $k \times 1$ vector). Notice the following:

$$\begin{aligned}E[g(X_i)y_i] &= E[g(X_i)X_i']\gamma + E[g(X_i)u_i]\\ E[g(X_i)y_i] &= E[g(X_i)X_i']\gamma \qquad\qquad (\text{ since } E[u_i|X_i] = 0)\\ \gamma &= E[g(X_i)X_i']^{+}E[g(X_i)y_i]\end{aligned}$$

Assuming $E[g(X_i)X_i']^{+}$ is full column rank with probability 1, we can write a least square estimator via the analogy principle:

$$\hat{\gamma} = (g(X)'X)^{-1}g(X)'y$$

where X and $g(X)$ are $n \times k$ matrices and y is a $n \times 1$ vector. For the particular problem given in the description, consider $X_i = [1\ x_i]'$ and $g(X_i) = [1\ x_i^3]'$. We can write the estimator as:

$$\begin{aligned}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} &= \left(\begin{bmatrix} n & \sum_i x_i\\ \sum_i x_i^3 & \sum_i x_i^4\end{bmatrix}\right)^{-1}\begin{bmatrix}\sum_i y_i\\ \sum_i x_i^3 y_i\end{bmatrix}\\ \begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} &= \frac{1}{m_n(x_i^4) - m_n(x_i^3)m_n(x_i)}\begin{bmatrix} m_n(x_i^4) & -m_n(x_i^3)\\ -m_n(x_i) & 1\end{bmatrix}\begin{bmatrix} m_n(y_i)\\ m_n(x_i^3 y_i)\end{bmatrix}\end{aligned}$$

where $m_n(k_i) = 1/n\sum_{i=1}^n k_i$. So $\hat{\beta} = \frac{m_n(x_i^3 y_i) - m_n(y_i)m_n(x_i)}{m_n(x_i^4) - m_n(x_i^3)m_n(x_i)}$ and $\hat{\alpha} = \frac{m_n(x_i^4)m_n(y_i) - m_n(x_i^3 y_i)m_n(x_i^3)}{m_n(x_i^4) - m_n(x_i^3)m_n(x_i)} = m_n(y_i) - m_n(x_i^3)\hat{\beta}$. Notice the $\hat{\beta}$ is different from the usual $\hat{\beta}_{standard} = \frac{Cov(x_i, y_i)}{Var(x_i)} = \frac{m_n(x_i y_i) - m_n(x_i)m_n(y_i)}{m_n(x_i^2) - m_n(x_i)m_n(x_i)}$. However, $\hat{\alpha}_{standard} = m_n(y_i) - m_n(x_i)\hat{\beta}_{standard}$, the same formulation as $\hat{\alpha}$.

Part (3)

First, let's note that a random variable y is mean independent of x if $E[y|x] = E[y]$.

Let's first show mean independence doesn't imply independence. An example that demonstrates this nicely that I have seen before is $x \sim U[1,2]$, $y \sim U[-x,x]$. Notice that $E[y|x=x] = 0 = \mathbb{E}[y]$, which gives us mean independence. However, x and y are clearly not independent as the distribution of y depends on x.

Let's show that independence leads to mean independence for discrete and continuous r.v.[1]:

- **Discrete:** Let x and y be independent random variables and consider the case where y is a discrete random variable. Then: $E[y|x] = \sum_y yP(y|x) = \sum_y yP(y) = E[y]$.
- **Continuous:** Let x and y be independent random variables and consider the case where y is a continuous random variable. Then: $E[y|x] = \int_{\mathcal{Y}} y f_{Y|X}(y|x)dy = \int_{\mathcal{Y}} y f_Y(y)dy = E[y]$.

where we are using the fact that $P(y|x) = P(y)$ for indepedent discrete r.v. and $f_{Y|X}(y|x) = f_Y(y)$ for independent continuous r.v..

Part (4)

(i)

By the law of iterated expectations,

$$\mathbb{E}[uh(x)] = \mathbb{E}_x\{\mathbb{E}[uh(x)|x]\} = \mathbb{E}_x\{h(x)\,\mathbb{E}[u|x]\} = 0 = \mathbb{E}[u]$$

(ii)

Note that $x \perp\!\!\!\perp u$ implies $x \perp\!\!\!\perp g(u)$ for all measurable functions $g(\cdot)$. Then

$$\begin{aligned}
\mathbb{E}_x[x]\,\mathbb{E}_u[g(u)] &= \mathbb{E}_x\{x\,\mathbb{E}_u[g(u)]\} \quad \text{since } \mathbb{E}_u[g(u)] \text{ is a constant} \\
&= \mathbb{E}_x\{x\,\mathbb{E}_{u|x}[g(u)|x]\} \quad \text{by independence} \\
&= \mathbb{E}_x\{\mathbb{E}_{u|x}[xg(u)|x]\} \quad \text{since we're conditioning on } x \\
&= \mathbb{E}_{x,u}[xg(u)] \quad \text{by the law of iterated expectations}
\end{aligned}$$

Part (5)

Consider $E[(y - \hat{y}(X))^2|X]$:

$$E[(y - \hat{y}(X))^2|X] = E[y^2|X] - 2E[y|X]\hat{y}(X) + \hat{y}(X)^2$$

1. Note that we can get a more general property using the independence of the sigma-algebras, but I am guessing this is not necessary for this class. If so, see Mahajan's lecture notes II, theorem 8.3 for the result that independence implies mean-independence.

considering $X = x$, $E[(y - \hat{y}(X))^2 | X = x]$ is a quadratic. Take the FOC:

$$\begin{aligned} 0 &= -2E[y|X = x] + 2\hat{y}(x) \\ E[y|X = x] &= \hat{y}(x) \end{aligned}$$

which implies that we should construct $\hat{y}$ such that $\hat{y}(x) = E[y|X = x]\ \forall x \in \mathcal{X}$. Or in other words an optimal $\hat{y}$ is $\hat{y}(X) = E[y|X]$. It is hard to comment exactly on its relation to u since we haven't provided any restrictions on u (such as mean zero). Consider $\epsilon = y - \hat{y}(X)$. Notice that $E[\epsilon|X] = E[y|X] - \hat{y}(X) = 0$. If we believe $\hat{y}(X)$ to be estimating f, then we can interpret $\epsilon \equiv u$.

Part (6)

For D to be a valid instrument, the lecture notes tell us that we need orthogonality and relevance. From the condition that $E[u|D] = 0$, it immediately follows that orthogonality is satisfied (i.e. $E[uD] = 0$). But, after discussing this problem with Lucy and Ethan, we agreed that the conditions given in the problem definition aren't sufficient to prove that $E[DX] \neq 0$[2]. This can be seen with a simple counter-example, but first notice the following equivalence:

$$\begin{aligned} E[DX] &= E[DE[X|D]] \\ &= E[X|D = 1]P(D = 1) \end{aligned}$$

Now consider the counter-example:

- $D \sim binom(p)$ for any $p \in (0, 1)$ and $X = a * (1 - D)$ for some constant $a \neq 0$.
- This implies that $E[X|D = 0] = a$ and $E[X|D = 1] = 0$, meaning $E[X|D] \neq E[X]$. So the conditions laid out in the problem statement are satisfied.
- This also implies $E[DX] = 0$, which would violate orthogonality as defined in the lecture notes.

the counter-example above seems like a good use of an instrument D, which would suggest that for binary variables, we should change our definition of our relevance definition. As Ethan suggested in his e-mail, $E[X|D = 0] \neq E[X|D = 1]$ might be sufficient. This condition follows immediately from $E[X|D] \neq E[X]$. Assuming we do have relevance and orthogonality, a consistent estimator of β would be $(D'X)^{-1}D'y$. This can also be written as $\frac{\sum_{i=1}^{n} D_i y_i}{\sum_{i=1}^{n} D_i X_i}$.

Part (7)

First, let's clarify the two models. OLS is an estimation method that is often used to estimate the β parameter in the model that has the following characteristics:

$$\text{(Classical model)} \quad y_i = X_i'\beta + \epsilon_i \quad E[\epsilon_i X_i] = 0$$

2. Note that the relevance condition in the slides states that $E[Z'X]$ needs to be full column rank, which implies that $E[ZX] \neq 0$ when Z and X are univariate.

whereas the IV estimator is often used if we have a model with the following characteristics:

$$\begin{aligned} \text{(Instrumental model)} \quad y_i &= X_i'\beta + \epsilon_i \\ X_i &= Z_i'\gamma + \eta_i \\ E[Z_i\epsilon_i] &= 0 \\ E[Z_i'X] &\text{ is full column rank} \end{aligned}$$

Figure 1 contains a drawing that is meant to represent the two causal diagrams. If we consider the drawing only in black, we can see a drawing that satisfies the characteristics of the classical model. If we look at the drawing in black plus red, we can see that the drawing no longer satisfies the classical model ($E[\epsilon_i X_i \neq 0]$[3]), however it does satisfy the characteristics of the instrumental model.

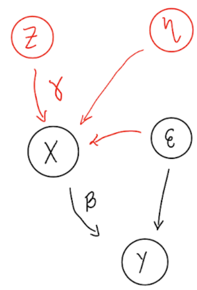

Figure 1: The figure is a hand drawing of a DAG. In black, one can see the DAG structure that is justified by the classical OLS model. In black + red, one can see endogeneity introduced for X and the addition of an instrument, creating a scenario where IV would be correctly specified.

The question also asks us to consider what would happen under mis-specification of the models. This can either be that the instrumental model is mis-specified or the classical model is mis-specified.

Let's first consider the case where the classical model is mis-specified but the instrumental model is correctly specified. This would imply that the $E[\epsilon_i X_i] \neq 0$ and so the red line between ϵ and X exists in Figure 1. If we know that the instrumental model is correctly specified[4], we can estimate both models and compare the parameter estimates for β.

Now consider the case where the instrumental model is mis-specified but the classical model is correctly specified. One of the two conditions might be violated. This would mean that our orthogonality or relevance conditions are violated. The second case can be assessed through looking at the relationship between Z and X, but there is not a good statistical way to assess the first violation.

We can also consider what happens if we estimate a mis-specified model. If we estimate the instrumental model when classical model is correct, we risk introducing bias through

3. Note that this figure doesn't eliminate the possibility that this condition holds since dependent r.v.s can be uncorrolated. But it eliminates the certainty (and likelihood) that this condition is met.

4. Testing for this is quite difficult if we do not have any control over ϵ. Since ϵ is usually an unobservable that as researchers we have no control over, we usually use qualitative arguments to justify this condition holding or not.

predicting X through an unrelated instrument. If we estimate the classical model when the instrumental model is correct, we introduce bias through violation of strict exogeneity- so called omitted variable bias.

2 Wright (1928)

In this section we are interested in the canonical demand and supply model. The general structure is the following:

- We have demand $q_D = D(p, u)$.
- We have supply $q_S = S(p, v)$.
- As denoted, u and v are random and we assume that they have distributions F_u and F_v. For simplicity, assume that the pdf's exist and are f_u and f_v respectively.
- We are interested in the market equilibrium where $D(p, u) = S(p, v)$, which is an implicit function where we can solve for $p^*(u, v)$ s.t. $D(p^*(u, v), u) = S(p^*(u, v), v) = q^*(u, v)$. Notice that since p^* is a function of two random variables, p^* also a r.v..

For the linear model that the notebook covers, we have the following model:

$$q_D = \alpha p + u \qquad q_S = \beta p + v \qquad q_D = q_S,$$

where (u, v) are distributed $(N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2))$ respectively. We can solve this problem when in equilibrium as:

$$\alpha p + u = \beta p + v \implies p^*(u, v) = (v - u)/(\alpha - \beta)$$

when plugging this back into our initial demand and supply equation:

$$q^*(u, v) = v(\alpha/(\alpha - \beta)) - u(\beta/(\alpha - \beta))$$

Part (1) - Control

In the setting where we are controlling price, we are no longer guaranteed equilibrium and price is no longer a random variable. However, demand still depends on u and as a result is random. As denoted in lecture notes, the expected demand will be:

$$E[D(p_0, u)] = \int D(p_0, u) f_u(u) du$$

for the linear model, this means:

$$E[D(p_0, u)] = E[\alpha p_0 + u] = \alpha p_0 + \mu_1$$

Part (2) - Condition

We want to consider the case where we observe $p = p_0$ as defined in the question and we believe that the market is in equilibrium (which we usually do). So:

$$E[q^*(u, v) | p^*(u, v) = p_0] = E[q^*(u, v) | q_D(p_0, u) = q_S(p_0, v)]$$

for the linear model:

$$p^*(u,v) = (v-u)/(\alpha-\beta) = p_0 \implies p_0(\alpha-\beta)+u = v$$

which implies demand is:

$$\begin{aligned} q^*(u,v)|(p^*(u,v)=p_0) &= q^*(u,v(u)) \\ &= v(u)(\alpha/(\alpha-\beta)) - u(\beta/(\alpha-\beta)) \\ &= [p_0(\alpha-\beta)+u](\alpha/(\alpha-\beta)) - u(\beta/(\alpha-\beta)) \\ &= p_0\alpha + u \end{aligned}$$

so:

$$E[q^*(u,v)|p^*(u,v)=p_0] = \alpha p_0 + \mu_1$$

Part (3) - Counterfactual

We want to consider the case where we observe (p_0, q_0). Let's first consider what we know. Note that we now know the following two equations:

$$\begin{aligned} &(1)\ D(p_0,u) = S(p_0,v) \\ &(2)\ q^*(u,v) = q_0 \end{aligned}$$

where (1) gives us an implicit function so that we can solve to get $v(u)$, as we did in the previous part. If in turn the event $q^*(u,v(u)) = q_0$ is a singleton, then for observing (p_0,q_0), we can pin down the (u_0,v_0) in our equations (no randomness). If we can pin down (u_0,v_0), we can further determine the change in demand due to the price change:

$$q_D(p_1,u_0) - q_D(p_0,u_0)$$

Let's go back to the linear model, where from the previous part we have that $q^*(u,v(u)) = p_0\alpha + u$. After observing q_0, we can solve this equation so that $u_0 = q_0 - p_0\alpha$. Which would in turn imply that $v_0 = q_0 - p_0\beta$. So the change in demand would be $\alpha(p_1 - p_0)$.

Counterfactual analysis assumes certain conditions, e.g. monotonicity, which ensure that the supply and demand curves only cross once at a given price point.

4. Weak Instruments

Please refer to the jupyter notebook Q4_notebook.ipynb for relevant code and examples

Part (1)

Complete - please see code/functions.py for function definition.

Part (2)

(a)

Complete - we wrote a class that performs Two-Stage Least Squares. See code/classes.py for class definition.

(b)

As seen in the Jupyter notebook, the 95% confidence interval of the Monte Carlo does adequately cover the 2SLS 95% confidence interval. Further, it is unbiased and rather precise. The t-test a test statistic of around 55, which returns a very small p-value.

(c)

We run Monte Carlos for 9 different values of π ranging from 0 to 1. For each pi, we take an average of our t-tests and p-values over the N Mote Carlo draws. The average size of the t statistic increases as π increases, indicating we are more confident that we can reject the null hypothesis as π increases.

Part (3)

(a)

As proposed in the problem, we create a dgp that depend on ℓ. From the dgp we created, adding increasingly weak instruments does not seem to harm the bias. We created three simulation exercises using the same dgp process that can be found in the notebook under "4.3.a".

For the first simulation, the coverage of the 95% confidence interval, interpreted as Upper CI - Lower CI, remains steady at around 0.125 for $L \in [1, 6]$. We also visually inspect the histograms and find nothing that caused us concern.

We then follow this up by a low "n" simulation. We set the data sample to 50 and calculate monte carlo averages for the RMSE, standard errors, and coverage. We calculate these values up to $\ell = 20$ and simulate this monte carlo 10,000 times. The second figure in this section plots the changes in these statistics for ℓ. We see an initial decrease in these statistics and then the values plateau.

Finally, to make sure these results hold for different "n", we replicate this exercise for $n \in \{50, 100, 500\}$. We find similar results.

(b)

From the analysis in part (a), it is not clear that there is an optimal number of instruments, as each of the tested values provide similarly low bias and coverage of a 95% confidence interval. But if we want to interpret "optimal" as the number of instruments after which we don't see any significant gains in RMSE, coverage and, s.e. estimates, it seems that the optimal number of instruments is around 5-6.

5. A simple approach to inference with weak instruments

Please refer to the jupyter notebook Q5_notebook.ipynb for relevant code and examples

Part (1)

Complete - please see code/functions.py for function definition.

Part (2)

A Monte Carlo exercise shows the bias is very small, and the estimator is quite precise - a mean of 1 and a variance of 0.001

Part (3)

The coverage of both the Hansen and the 2SLS estimators under a Monte Carlo setting is about 0.12 when $\pi = 1$, indicating both estimators are equally precise when $\pi = 1$

Part (4)

We calculate coverage, interpreted as (upper CI - lower CI) for Monte Carlos of 10 values of π ranging from 0 to 1 for both the Hansen estimator and 2SLS. At low values of π, the coverage of the hansen estimator blows up much larger than 2SLS. However by $\pi \geq 0.25$, they are nearly identical in coverage.

Part (5)

As we add increasingly weak instruments, our coverage decreases slightly implying a slightly more precise estimator. Comparing this to 2SLS, whose coverage was unaffected by adding sequentially weaker estimators.