# Econ 21020 - Problem Set 5 Solutions

## Problem 1

You want to estimate the proportion of UChicago students who have ever cheated. Call the true value of this proportion $\theta$. However, you think if you just ask students this question, they might lie. Instead, you collect an iid sample of $n$ students, where you give each student the instructions:

1. Flip a fair coin (50/50) secretly.
2. If the coin comes up heads, answer the question, "Have you every cheated?" If the coin comes up tails, give the response "Yes."

Under this procedure, you assume that everyone will answer honestly if they get a heads. Let $X_i$ denote the response of the $i$th student, where

$$X_i = \begin{cases} 1 \text{ if they say "Yes"} \\ 0 \text{ if they say "No"} \end{cases}$$

However, you do not observe the outcome of the coin toss for each student.

(a) As a function of $\theta$, what is $P\{X_i = 1\}$?

SOLUTION: We could go just by intuition here - there's a 1/2 chance of heads, at which point there's a $\theta$ chance of saying yes, aka $X_i = 1$, and there's a 1/2 chance of tails, at which point $X_i = 1$:

$$P(X_i = 1) = \frac{\theta}{2} + \frac{1}{2} = \frac{\theta + 1}{2}$$

We could also be more formal by defining a new variable $Z_i$ where 1 indicates a heads and 0 tails, say that and then say that

$$\begin{aligned} P(X_i = 1) &= P(X_i = 1, Z_i = 1) + P(X_i = 1, Z_i = 0) \\ &= P(X_i = 1|Z_i = 1)P(Z_i = 1) + P(X_i = 1|Z_i = 0)P(Z_i = 0) \\ &= \theta * \frac{1}{2} + 1 * \frac{1}{2} \\ &= \frac{\theta + 1}{2} \end{aligned}$$

(b) Show that $\hat{\theta}_n = \frac{2}{n}\sum_{i=1}^{n} X_i - 1$ is a consistent estimator for $\theta$.

SOLUTION: We're showing consistency, and our estimator involves a sample mean, so let's start by applying the WLLN. We have that that the $X_i$ are iid. We also know that $E[X_i^2] < \infty$ because $X^2 \leq 1$. Thus, we can apply the WLLN to say:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E[X]$$

Generally, $E[\mathbb{1}\{Y \in A\}] = P(Y \in A)$, and, for $X \in \{0,1\}$, $X = \mathbb{1}\{X = 1\}$ (they both only take on the values 0 and 1, and take on those values at the same time), so we can further say:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E[X] = P(X = 1)$$

From part a), we know that $\theta = 2P(X = 1) - 1$, which is a continuous function of $P(X = 1)$. Thus, we can apply the CMT to say:

$$\hat{\theta}_n = \frac{2}{n}\sum_{i=1}^{n} X_i - 1 \xrightarrow{p} 2P(X = 1) - 1 = \theta$$

so the estimator is consistent by definition.

(c) Find the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ (the distribution as $n \to \infty$).

SOLUTION: We're finding a limiting distribution, so let's apply the CLT. The CLT applies to sample means, so let's apply it to the sample mean within $\hat{\theta}_n$. We still have an iid sample and know that $E[X] < \infty$, so by the CLT:

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - P(X = 1)) \xrightarrow{d} N(0, Var(X))$$

We can stop down here to see that, as $X$ is a Bernoulli random variable,

$$Var(X) = P(X = 1)(1 - P(X = 1))$$

In order to now get to the estimator, we need to subtract a 1 from $\frac{1}{n}\sum_{i=1}^{n} X_i$ and multiply it by 2. We can get the multiplication by 2, using Slutsky (2 trivially converges in probability to 2):

$$2\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - P(X = 1)) \xrightarrow{d} 2N(0, Var(X))$$

$$\sqrt{n}(\frac{2}{n}\sum_{i=1}^{n} X_i - 2P(X = 1)) \xrightarrow{d} N(0, 4Var(X))$$

The final bit follows because a random variable multiplied by any number will have it's variance multiplied by the square of that number. Now we can bring in the negative 1 by adding and subtracting a 1 on the left hand side:

$$\begin{aligned}\sqrt{n}(\frac{2}{n}\sum_{i=1}^{n} X_i - 1 + 1 - 2P(X=1)) &= \sqrt{n}(\frac{2}{n}\sum_{i=1}^{n} X_i - 1 - (2P(X=1) - 1)) \\ &= \sqrt{n}(\hat{\theta}_n - \theta) \\ &\xrightarrow{d} N(0, 4Var(X))\end{aligned}$$

the second line finishes by using the definition of $\hat{\theta}_n$ and the expression for $\theta$ we found in part a).

(d) Propose an estimator, $\hat{\sigma}$, such that

$$\frac{1}{\sqrt{\hat{\sigma}}}\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1)$$

Show that this convergence in distribution takes place.

SOLUTION: We can normalize any mean-zero, normally distributed random variable by dividing by it's standard deviation. We see that our normally distribution random variable has the variance

$$4Var(X) = 4\hat{\theta}_n P(X=1)(1 - P(X=1))$$

as discussed in the previous part. We already know that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} P(X=1)$$

and we know that $4A(1-A)$ is continuous for any $A$, so, by the CMT,

$$4\frac{1}{n}\sum_{i=1}^{n} X_i(1 - \frac{1}{n}\sum_{i=1}^{n} X_i) \xrightarrow{p} 4P(X=1)(1 - P(X=1)) = 4Var(X)$$

We can go ahead and apply CMT again to say that:

$$\sqrt{4\frac{1}{n}\sum_{i=1}^{n} X_i(1 - \frac{1}{n}\sum_{i=1}^{n} X_i)} \xrightarrow{p} \sqrt{4P(X=1)(1 - P(X=1))} = \sqrt{4Var(X)}$$

as the square root is a continuous function. Thus, we have a consistent estimator for the standard deviation of our limiting distribution. In order to divide by it, we can use Slutsky's Lemma, as we have an estimator of the standard deviation that is converging in probability and $\sqrt{n}(\hat{\theta}_n - \theta)$ which is converging in distribution. By Slutsky:

$$\frac{\sqrt{n}}{\sqrt{4\frac{1}{n}\sum_{i=1}^n X_i(1-\frac{1}{n}\sum_{i=1}^n X_i)}}(\hat{\theta}_n - \theta) \xrightarrow{d} \frac{1}{\sqrt{4Var(X)}}N(0, 4Var(X))$$
$$\xrightarrow{d} N(0,1)$$

this is legal, so long as $\sqrt{4Var(X)} > 0$, which is the case if $Var(X) = P(X=1)(1-P(X=1)) > 0$, which occurs as long as we don't have $\theta = 1$ (aka it is not the case that 100% of students are cheating).

## Problem 2

In the multivariate case of linear regression, we brushed over the idea of homoskedasticity, and went right to heteroskedasticity robust inference. However, we can define homoskedasticity analagously in the multivariate case: homoeskedasticity holds if $E[U|X] = 0$ and $Var(U|X) = Var(U)$, where $X$ is a $(k+1) \times 1$ random vector.

(a) Show that

$$\Sigma = E[XX']^{-1}Var(XU)E[XX']^{-1}$$

is equal to

$$\Sigma^{Ho} = E[XX']^{-1}Var(U)$$

if $U$ is homoskedastic. (Hints: Start by working with $Var(XU)$. The definitional of the conditional variance may be useful. This will in general look similar to an analagous result in the univariate case.)

SOLUTION: As suggested, we'll start by working with $Var(XU)$:

$$\begin{aligned} Var(XU) &= E[(XU - E[XU])(XU - E[XU])'] \\ &= E[(XU)(XU)'] \end{aligned}$$

The following holds because, in a regression context, we need to assume $E[XU] = 0$. (Alternatively, working solely from the definition of homoskedasticity, we could say:

$$\begin{aligned} E[UX] &= E[E[UX|X]] && \text{(LIE)} \\ &= E[XE[U|X]] \\ &= E[0] = 0 \end{aligned}$$

getting the same thing.) Proceeding from there, we can take advantage of the fact that $U$ is univariate, so it can "commute around" within matrix multiplication:

$$\begin{aligned} Var(XU) &= E[(XU)(XU)'] \\ &= E[U^2 XX'] \\ &= E[E[U^2 XX'|X]] && \text{(LIE)} \\ &= E[E[U^2|X]XX'] \end{aligned}$$

Now we can bring in the other hint and take a look a the definition of conditional variance:

$$\begin{aligned} Var(U|X) &= E[U^2|X] - E[U|X]^2 \\ \Rightarrow E[U^2|X] &= Var(U|X) + E[U|X]^2 \\ E[U^2|X] &= Var(U) \end{aligned}$$

where the last line follows by applying both of the components of the definition of homoskedasticity. We can plug this into our working expression for $Var(XU)$:

$$\begin{aligned} Var(XU) &= E[E[U^2|X]XX'] \\ &= E[Var(U)XX'] \\ &= Var(U)E[XX'] \end{aligned}$$

Now we can at last go ahead and plug into the expression for $\Sigma$ to finish:

$$\begin{aligned} \Sigma &= E[XX']^{-1}Var(XU)E[XX']^{-1} \\ &= E[XX']^{-1}Var(U)E[XX']E[XX']^{-1} \\ &= E[XX']^{-1}Var(U) \end{aligned}$$

(b) Show that $Var(U) = E[U^2]$, (and so $\Sigma^{Ho} = E[XX']^{-1}E[U^2]$) under any interpretation of linear regression.

SOLUTION: This follows from the definition of variance:

$$\begin{aligned} Var(U) &= E[U^2] - E[U]^2 \\ &= E[U^2] \end{aligned}$$

For any regression model we can claim that $E[U] = 0$ to get the above. (Alternatively, working directly from the definition of homoskedasticity:

$$\begin{aligned} E[U] &= E[E[U|X]] && \text{(LIE)} \\ &= E[0] = 0 \end{aligned}$$

instead.)

## Problem 3

In the case of multivariate linear regression, we can test more types of hypotheses than we did in the univariate case. Let's look at one other type: testing a hypothesis that two subcomponents of $\beta$ are equal. That is, consider a regression equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

and the hypotheses $H_0 : \beta_1 = \beta_2$, $H_1 : \beta_1 \neq \beta_2$. These are equivalent to the hypotheses $H_0 : r'\beta = 0$, $H_0 : r'\beta \neq 0$ for the $3 \times 1$ vector:

$$r = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

(a) Find the limiting distribution of:

$$r'\sqrt{n}(\hat{\beta} - \beta)$$

Represent the variance of the limiting distribution in terms of $r$ and $\Sigma$, where $\Sigma$ is the typical variance of the limiting distribution of $\sqrt{n}(\hat{\beta} - \beta)$,

$$\Sigma = E[XX']^{-1} Var(XU) E[XX']^{-1}$$

SOLUTION: We know that,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \Sigma)$$

So by Slutsky's lemma:

$$\begin{aligned} r'\sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} r'N(0, \Sigma) \\ &\xrightarrow{d} N(0, r'\Sigma r) \end{aligned}$$

The first line follows from Slutsky's Lemma (this might seem odd, but for a constant like, $r$, $r \xrightarrow{p} r$ trivially. Thus, Slutsky's Lemma technically allows us to bring constants into sequences that are converging in distribution). The second line follows from properties of multivariate normal distributions.

(b) What are the dimensions of the variance of the limiting distribution of $r'\sqrt{n}(\hat{\beta} - \beta)$?

SOLUTION: The variance is $r'\Sigma r$. This is something that is $1 \times 3$ times something that is $3 \times 3$ times something that is $3 \times 1$. Remember that with matrix multiplication, the "inner" dimensions have to match, and the product takes on the "outer" dimensions. Thus, the whole thing will end being $1 \times 1$, aka univariate.

(c) Propose an estimator $\hat{\sigma}$ such that

$$\frac{1}{\sqrt{\hat{\sigma}}} r' \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, 1)$$

Show that this convergence in distribution takes place. (Hint: Take a look at the version of inference on multivariate linear regression that we did in class. You may make use of the result that $\hat{\Sigma}$, as defined in lecture, is a consistent estimator for $\Sigma$.)

SOLUTION: We can get the above if $\hat{\sigma}$ is a consistent estimator for $\sqrt{r'\Sigma r}$. We know that $\hat{\Sigma} \xrightarrow{p} \Sigma$, so by direct application of the CMT:

$$\sqrt{r'\hat{\Sigma} r} \xrightarrow{p} \sqrt{r'\Sigma r}$$

as matrix multiplication and taking square roots are both continuous functions. Finally, we can apply Slutsky (assuming that $r'\Sigma r \neq 0$), to say:

$$\begin{aligned} \frac{1}{\sqrt{r'\hat{\Sigma} r}} \sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} \frac{1}{\sqrt{r'\Sigma r}} N(0, r'\Sigma r) \\ &\xrightarrow{d} N(0, 1) \end{aligned}$$

where the last line follows from properties of (univariate) normal distributions.

(d) Think of real-world $Y$, $X_1$, and $X_2$ for which this type of hypothesis test may be interesting.

SOLUTION: Basically we want to think of any case where we might want to compare two effects to see if they're identical or not. Perhaps we want to compare two types of treatments, where one is more expensive than the other, so we want to know if they produce differential effects (if not we'd just prefer the cheaper option). Then, we could have $Y$ be the outcome of interest and $X_1$ and $X_2$ be indicators for each of the two treatments. Specific example: maybe we're comparing two after-school programs, indicated by $X_1$ and $X_2$ respectively, on an outcome like HS graduation. If we interpret the regression causally, a hypothesis test of $H_0 : \beta_1 = \beta_2$ is testing whether the two programs have an identical effect. If they don't have an identical effect, we'll presumable prefer the one that is more effective.

# Problem 4

This question develops another example of a LATE, based on the paper Angrist (1990).

(a) Consider a causal model, of the effect of military service on (post-service) wages. Specifically,:

$$W = \beta_0 + \beta_1 M + U$$

where

$$W = wage$$
$$M = \begin{cases} 1 \text{ if served in the military} \\ 0 \text{ if not} \end{cases}$$

Why might it be the case that $E[MU] \neq 0$, where $U$ is defined in the causal model sense?

SOLUTION: We'll have $E[MU] \neq 0$ if causal determinants of wage, other than military service, are correlated with military service. There's a wide host of reasons why this might be the case. For one thing, the military is overwhelmingly male and males tend to earn higher wages. This will induce a correlation between military service and $U$. There are many other examples to think of.

(b) Consider now an instrumental variables approach, using the military draft implemented by the US government during the Vietnam War as an instrument. For the sake of this example, assume the draft works very "simply" - every single US man aged 19-26 is entered into the draft and a subset are drafted.[1] Those who are drafted are called up to service by the government, under threat of legal action. Define the new variable, $D$:

$$D = \begin{cases} 1 \text{ if drafted} \\ 0 \text{ if not} \end{cases}$$

The LATE interpretation of IV requires three assumptions:

(a) $(W_1, W_0, M_1, M_0) \perp D$ (implies instrument exogeneity)
(b) $M_1 \neq M_0$ sometimes (analogous to instrument relevance)
(c) $M_1 \geq M_0$ always - called Monotonicity

where $W_1$ and $W_0$ are the potential outcomes corresponding to the two values of $M$ and $M_1$ and $M_0$ are the potential treatments corresponding to the two values of $D$. Evaluate each of the three LATE assumptions in this context (are they reasonable to assume? Why or why not?).

SOLUTION: If we assume that men are drafted at random, then $(W_1, W_0, M_1, M_0) \perp D$ would be reasonable. The other characteristics of men that are drafted, which determine the potential outcomes and potential treatments, should then be random with respect to draft status. (If we think that men are drafted non-randomly we could reject this assumption).

$M_1 \neq M_0$ sometimes requires that the potential treatment will sometimes vary with the instrument. This would suggest that there are at least some

[1] From now on, we will consider the "population" to be US men aged 19-26 at the time of the Vietnam War draft.

men for whom military service status depends on draft service. There are presumably very many men who would not have served in the military if not drafted but did serve after being drafted - for any such man, we'd have $M_1 = 1$ and $M_0 = 0$, so $M_1 \neq M_0$.

The final assumption, monotonicity says that the instrument only shifts people in one direction. In this context it is saying that there are no men who *would* serve in the military if not drafted but *would not* serve in the military if drafted. It seems generally reasonable to think that someone who would have voluntarily served in the military would not refuse if they were compelled to do so, so monotonicity is likely sound.

(c) We can split the population into three groups:

- Always-takers: People for whom $M_1 = 1,\ M_0 = 1$
- Never-takers: People for whom $M_1 = 0,\ M_0 = 0$
- Compliers: People for whom $M_1 = 1,\ M_0 = 0$

Interpret in words who the members of each group are.

SOLUTION: The always-takers would sign up for the military regardless of whether or not they are drafted. The never-takers would not sign up for the military in any case (so in our context this means that if they were drafted, never-takers fled the country, suffered the legal penalty of draft avoidance, or otherwise somehow got out of it). The compliers are people who would not serve in the military if not drafted but would do so if they were drafted.

(d) Define the LATE for this regression. Interpret it in words. Is this LATE interesting (this last question is more or less wholly subjective - feel free to argue either way, demonstrating your knowledge of what a LATE is)?

SOLUTION: In this context, the LATE will be:

$$E[W_1 - W_0 | M_1 > M_0]$$

This is the average difference in wage between the world in which the person served and the world in which the person did not serve, averaged across the portion of the population who would only serve in the military if legally forced to. Whether this is interesting or not is an interesting question. From a direct policy perspective, it seems unlikely that the US would ever start forcing all men to serve in the military, so knowing the effect on wage of compelling un-willing people to join the military might not be very policy relevant. However, it could be interesting from an intellectual perspective to know what the effect on such people would be. Moreover, maybe we think the compliers (reluctant servicemen) are not so different from the always-takers (the willing servicemen)? We might then be willing to extrapolate the LATE to people who actually serve in the military during peace time, which might be useful information.

## Problem 6

We now conclude our discussion of the data from Angrist and Krueger (1999), following fairly close the identification strategy that they use in their paper. We will again refer to $X$ as years of education and $Y$ as log wage. We will continue to assume that $E[X^4], E[Y^4] < \infty$.

Consider quarter of birth as an instrument for number of years of education completed. The idea of this instrument is thus - students are legally required to attend school until a certain age. Given the cutoffs for students to be assigned to grades in US schools, students born earlier in the year tend to be older than their peers are, during each grade. Therefore, students born earlier in the year will tend to, legally, be allowed to drop out of high school in earlier grades than their peers. In this way, quarter of birth will affect years of education completed while being, potentially, independent of other causal determinants of wage. Consider the following graph, reproduced from Angrist and Kreuger (1999):

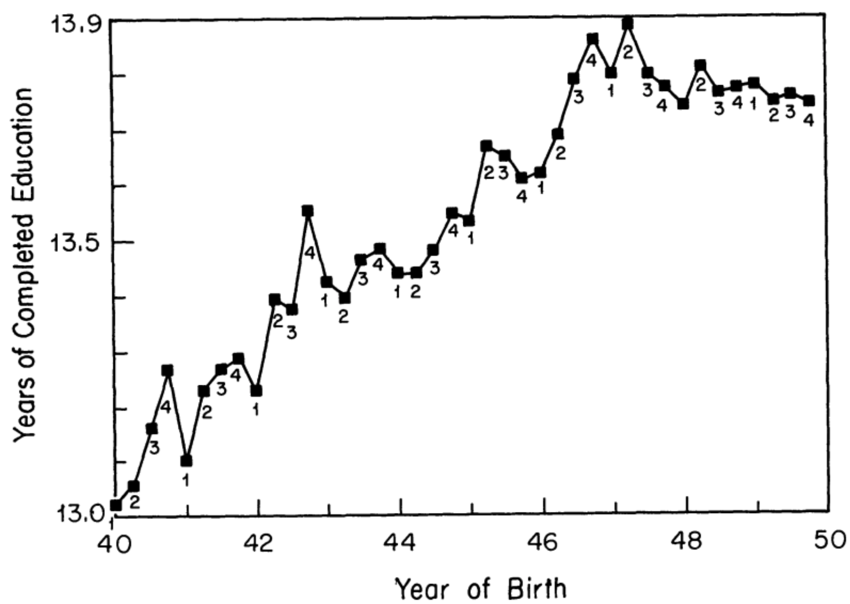


**FIGURE II**
**Years of Education and Season of Birth**
**1980 Census**
***Note.* Quarter of birth is listed below each observation.**

For convenience, let's define our instrument, $Z$, as:

$$Z = \begin{cases} 1 \text{ if born in quarter 1} \\ 0 \text{ if born in quarters 2-4} \end{cases}$$

(a) For a valid IV, we need to satisfy two assumptions: instrument exogeneity and instrument relevance. Do you think these will be valid in this context? Why or why not?

SOLUTION: Instrument relevance seems fairly straightforward - this says that $Cov(Z, X) \neq 0$. From the graph, there's a fairly clear pattern that people born earlier in the year (particularly in the 1st quarter) get fewer years of education. An association between the two variables will then suggest that $Cov(Z, X) \neq 0$, so relevance would be satisfied.

Instrument exogeneity is more interesting. This says that $E[ZU] = 0$, so quarter of birth is unrelated to other causal determinants of wage, other than years of schooling. Generally, quarter of birth seems fairly "random" - there might not be reason to think that families with different characteristics would have kids at different types of year, which would suggest that the exogeneity assumption is valid. However, there could be some minor concerns. In a grade, students who are older than their classmates tend to perform (slightly) better academically (presumably because their slight age advantage gives them a leg up, particularly in earlier grades, where a month or two matters). If this is indeed the case, it would suggest that quarter of birth might be correlated with, for instance, GPA, which might be a component of $U$.

(b) Calculate the IV estimand for the equation:

$$Y = \beta_0 + \beta_1 X + U$$

using $Z$ as an instrument, using your preferred software. Interpret the output.

(c) Suppose we considered a binary version of our schooling variable,

$$X' = \begin{cases} 1 \text{ if graduated HS} \\ 0 \text{ if not} \end{cases}$$

Then, if we assumed our LATE assumptions are true, we could interpret the paramter $\beta_1'$ from

$$Y = \beta_0' + \beta_1' X' + U'$$

using $Z$ as an instrument, as a LATE. Write down an expression for this LATE and interpret it in words.

SOLUTION: This LATE would look like:

$$E[Y_1 - Y_0 | X_1 < X_0]$$

where $Y_1$ and $Y_0$ are the potential outcomes telling us the wage for a student if they did vs. did not graduate HS and $X_1$ and $X_0$ are the potential treatments telling us whether or not a student graduated HS in the hypothetical world where they are born in the 1st quarter of the year vs the hypothetical world where they are born in a different quarter (I said $X_1 < X_0$ because students born earlier tend to get *less* schooling than others. The monotonicity assumption works in either direction, if $X_1 \leq X_0$ always or $X_1 \geq X_0$ always - the "always" is the important part). Let's think about the complier group first. These are kids who would only graduate HS if they were born outside the 1st quarter. Then, the LATE is the average effect of HS graduation on wages for this segment of the population. Is this a valuable LATE? It's slightly unclear - we might think of kids who will only graduate if they are born later in the year as kids who are on the "margin" of graduating HS. If we get a positive LATE, this might suggest that trying to encourage graduation among kids who are on the edge of graduating or not would be helpful to those kids.