Empirical Analysis III Mogstad Problem Set 3

Chase Abram, Tom Hierons, Jeanne Sorin

July 1, 2020

Problem 1

(a) Let A denote the event of always takers ($D_0 = 1, D_1 = 1$), and C denote the event of a compliers ($D_0 = 0, D_1 = 1$).

$$\begin{aligned}
ATT &= \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid D_Z = 1] \\
&= \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid D_0 = 0, D_1 = 1, D_Z = 1] \cdot \mathbb{P}[D_0 = 0, D_1 = 1 \mid D_Z = 1] \\
&\quad + \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid D_0 = 1, D_1 = 1, D_Z = 1] \cdot \mathbb{P}[D_0 = 1, D_1 = 1, \mid D_Z = 1] \\
&\quad + \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid D_0 = 1, D_1 = 0, D_Z = 1] \cdot \mathbb{P}[D_0 = 1, D_1 = 0, \mid D_Z = 1] \\
&\hspace{20em} \text{(Law of Tot. Pr.)} \\
&= \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid D_0 = 0, D_1 = 1, D_Z = 1] \cdot \mathbb{P}[D_0 = 0, D_1 = 1 \mid D_Z = 1] \\
&\quad + \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid D_0 = 1, D_1 = 1, D_Z = 1] \cdot \mathbb{P}[D_0 = 1, D_1 = 1, \mid D_Z = 1] \\
&\hspace{20em} \text{(Monotonicity)} \\
&= \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid C, D_Z = 1] \cdot \mathbb{P}[C \mid D_Z = 1] \\
&\quad + \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid A, D_Z = 1] \cdot \mathbb{P}[A, \mid D_Z = 1] \\
&= [ATT \mid A] \cdot \mathbb{P}[A \mid D_Z = 1] + [ATT \mid C] \cdot \mathbb{P}[C \mid D_Z = 1]
\end{aligned}$$

The last line denotes that the total ATT is the weighted average of the ATTs for the always-taker and complier sub-populations, where the weights are given by the probability of being in either population, given that treatment occurs.

(b) We can recall

$$\begin{aligned}
LATE &= \mathbb{E}[Y_{1,Z} - Y_{0,Z} \mid C] \\
ATE &= \mathbb{E}[Y_{1,Z} - Y_{0,Z}]
\end{aligned}$$

Then it becomes plain that the above effects are equal when the compliance is independent of treatment effect, i.e. if A is a Borel set, then $\mathbb{P}[(Y_{1,Z} - Y_{0,Z} \in A) \cap C] = \mathbb{P}[Y_{1,Z} - Y_{0,Z} \in A] \cdot \mathbb{P}[C]$. Note that full compliance is a special case of this condition.

Economics: Is this likely to be true? Doubtful, we might think that people are more likely to take if they will see a higher treatment effect. However, if we think full compliance is probable (i.e. everyone who wins the lottery (Z), chooses to collect lottery earning (D), and then we see how that affects their bequest at death (Y)), then this would be appropriate.

Problem 2

(a) By "identify," we here simply mean that, in population terms, an expression can be recovered by knowing the distribution F of (Y_D, X, D_Z, Z) (along with the given assumptions). It is important to be clear that this is different from knowing the distribution G of $(Y_1, Y_0, X, D_1, D_0, Z)$, in which case all these questions would be trivially known. It is also important to be clear that we are saying nothing about estimation techniques.

Let a be always takers, n be never takers, and d be defiers.

$$
\begin{aligned}
LATE(x) &\equiv \mathbb{E}[Y_1 - Y_0 \mid T = c, X = x] \\
&= \mathbb{E}[Y_{D_1} - Y_{D_0} \mid T = c, X = x] && \text{(Def. of } T = c\text{)} \\
&= \mathbb{P}[T = c \mid X = x]^{-1} \cdot \Big[\mathbb{E}[Y_{D_1} - Y_{D_0} \mid X = x] - \mathbb{E}[Y_{D_1} - Y_{D_0} \mid T = a, X = x] \cdot \mathbb{P}[T = a \mid X = x] \\
&\quad - \mathbb{E}[Y_{D_1} - Y_{D_0} \mid T = n, X = x] \cdot \mathbb{P}[T = n, X = x] \\
&\quad - \mathbb{E}[Y_{D_1} - Y_{D_0} \mid T = d, X = x] \cdot \mathbb{P}[T = d, X = x]\Big] \\
& && \text{(Law of Tot. Pr. (rearranged))} \\
&= \mathbb{P}[T = c \mid X = x]^{-1} \cdot \Big[\mathbb{E}[Y_{D_1} - Y_{D_0} \mid X = x] - 0 \cdot \mathbb{P}[T = a \mid X = x] \\
&\quad - 0 \cdot \mathbb{P}[T = n, X = x] \\
&\quad - \mathbb{E}[Y_{D_1} - Y_{D_0} \mid T = d, X = x] \cdot \mathbb{P}[T = d, X = x]\Big] && \text{(Def. of } a,\ n\text{)} \\
&= \mathbb{P}[T = c \mid X = x]^{-1} \cdot \Big[\mathbb{E}[Y_{D_1} - Y_{D_0} \mid X = x] - \mathbb{E}[Y_{D_1} - Y_{D_0} \mid T = d, X = x] \cdot 0\Big] \\
& && \text{(Monotonicity)} \\
&= \mathbb{P}[T = c \mid X = x]^{-1} \cdot \mathbb{E}[Y_{D_1} - Y_{D_0} \mid X = x]
\end{aligned}
$$

The expectation is known under our knowledge of F[1]. The probability term can be found as

$$
\begin{aligned}
\mathbb{P}[T = c \mid X = x] &= 1 - \mathbb{P}[T = a \mid X = x] - \mathbb{P}[T = n \mid X = x] - \mathbb{P}[T = d \mid X = x] \\
&= 1 - \mathbb{P}[T = a \mid X = x] - \mathbb{P}[T = n \mid X = x] && \text{(Monotonicity)} \\
&= 1 - \mathbb{P}[D_1 = 1, D_0 = 1 \mid X = x] - \mathbb{P}[D_1 = 0, D_0 = 0 \mid X = x] \\
&= 1 - \mathbb{P}[D_0 = 1 \mid X = x] - \mathbb{P}[D_1 = 0 \mid X = x] && \text{(Monotonicity)} \\
&= 1 - \mathbb{P}[D_Z = 1 \mid X = x, Z = 0] - \mathbb{P}[D_Z = 0 \mid X = x, Z = 1] && \text{(Exogeneity)} \\
&= \mathbb{P}[D_Z = 1 \mid X = x, Z = 1] - \mathbb{P}[D_Z = 1 \mid X = x, Z = 0] && \text{(Law of Tot. Pr.)} \\
&= \mathbb{E}[D_Z \mid X = x, Z = 1] - \mathbb{E}[D_Z \mid X = x, Z = 0]
\end{aligned}
$$

Both of the terms in the final line above are known under F. Therefore $LATE(x)$ is identified[2]. Note that relevance buys us that the penultimate line above is not zero, which is needed to ensure the original term is actually invertible, as we require in the conclusion of the first calculation. Furthermore, though we did not explicitly state it at each step, the overlap condition ensures that it was legal to condition on $Z = 0$ or $Z = 1$ anytime we needed to.

[1] More intuitively, we can find the average of the Ys such that $Z = 1$ and the Ys such that $Z = 0$. This is exactly what the expectation term is differencing.

[2] For all x in the support of X.

(b) Let $f_.$ denote a respective density.

$$
\begin{aligned}
\mathbb{E}[Y_1 - Y_0 \mid T = c] &= \mathbb{E}[\mathbb{E}[Y_1 - Y_0, \mid T = c, X = x] \mid T = c] && \text{(LIE)} \\
&= \mathbb{E}[LATE(X) \mid T = c] \\
&= \int_X LATE(x) f_{X|T=c}(x)\mathrm{d}x && \text{(Def. of } \mathbb{E}\text{)} \\
&= \int_X LATE(x) \frac{f_{X,T=c}(x, T = c)}{\mathbb{P}[T = c]}\mathrm{d}x && \text{(Bayes)} \\
&= \int_X LATE(x) \frac{\mathbb{P}[T = c \mid x]}{\mathbb{P}[T = c]} f_X(x)\mathrm{d}x && \text{(Bayes)} \\
&= \mathbb{E}\Big[\frac{LATE(X)\mathbb{P}[T = c \mid X]}{\mathbb{P}[T = c]}\Big]
\end{aligned}
$$

(c) The TSLS estimator projects[3] the explanatory part of Z on D into $\hat{D}$, then projects this estimate into Y space. We also need to consider X as part of each stage. So we may define the following

$$
\begin{aligned}
\beta_{TSLS} &= \frac{\mathrm{cov}(Y, \tilde{D})}{\mathrm{var}(\tilde{D})} \\
\tilde{D} &= \hat{D} - \mathbb{E}[\hat{D} \mid X] \\
\hat{D} &= \mathbb{E}[Z^2 \mid X]^{-1}\mathbb{E}[Z(D - \mathbb{E}[D \mid X]) \mid X]Z + \mathbb{E}[D \mid X]
\end{aligned}
$$

The subtracting of $\mathbb{E}[D \mid X]$ accounts for the fact that, since we are conditioning on X, there may be a part of the projection that is dependent upon X[4]. Now can consider the numerator and denominator of β_{TSLS} separately.

$$
\begin{aligned}
\mathrm{cov}(Y, \tilde{D}) &= \mathbb{E}[Y\tilde{D}] - \mathbb{E}[Y]\mathbb{E}[\tilde{D}] \\
&= \mathbb{E}[Y\tilde{D}] && (\mathbb{E}[\tilde{D}] = 0 \text{ by LIE}) \\
&= \mathbb{E}[Y(\hat{D} - \mathbb{E}[\hat{D} \mid X])] \\
&= \mathbb{E}[\mathbb{E}[Y(\hat{D} - \mathbb{E}[\hat{D} \mid X]) \mid Z, X]] \\
&= \mathbb{E}[\mathbb{E}[Y \mid X, Z](\hat{D} - \mathbb{E}[\hat{D} \mid X])] && (\tilde{D}, \mathbb{E}[\hat{D} \mid X] \in \sigma(X, Z))
\end{aligned}
$$

This expression is going to become unwieldy, but we can try to minimize this by considering the first term inside first.

[3]Generally, linear projection is *not* equivalent to condition expecation. However, here we assume the true model is linear, so we may use the BLP and CEF interchangeably.

[4]Problem Set 2 Problem 1 details this idea.

$$\begin{aligned}
\mathbb{E}[Y \mid X, Z] &= \mathbb{E}[Y \mid X, Z=0] + Z\big(\mathbb{E}[Y \mid X, Z=1] - \mathbb{E}[Y \mid X, Z=0]\big) && (Z \text{ binary}) \\
&= \mathbb{E}[Y \mid X, Z=0] + Z\frac{\mathbb{E}[Y \mid X, Z=1] - \mathbb{E}[Y \mid X, Z=0]}{\mathbb{E}[D \mid X, Z=1] - \mathbb{E}[D \mid X, Z=0]}\big(\mathbb{E}[D \mid X, Z=1] - \mathbb{E}[D \mid X, Z=0]\big) \\
&= \mathbb{E}[Y \mid X, Z=0] + ZLATE(X)\big(\mathbb{E}[D \mid X, Z=1] - \mathbb{E}[D \mid X, Z=0]\big) \\
& && (LATE(X) = ATE \mid c, X)
\end{aligned}$$

Now we return

$$\begin{aligned}
\operatorname{cov}(Y, \tilde{D}) &= \mathbb{E}\Big[\mathbb{E}[Y \mid X, Z=0](\hat{D} - \mathbb{E}[\hat{D} \mid X) \\
&\qquad + ZLATE(X)\big(\mathbb{E}[D \mid X, Z=1] - \mathbb{E}[D \mid X, Z=0]\big)(\hat{D} - \mathbb{E}[\hat{D} \mid X])\Big] \\
&= \mathbb{E}\Big[ZLATE(X)\big(\mathbb{E}[D \mid X, Z=1] - \mathbb{E}[D \mid X, Z=0]\big)(\hat{D} - \mathbb{E}[\hat{D} \mid X])\Big] && \text{(LIE)} \\
&= \mathbb{E}\left[LATE(X)\mathbb{E}\Big[Z\big(\mathbb{E}[D \mid X, Z=1] - \mathbb{E}[D \mid X, Z=0]\big)(\hat{D} - \mathbb{E}[\hat{D} \mid X]) \mid X\Big]\right] \\
& && \text{(LIE)} \\
&= \mathbb{E}\left[LATE(X)\mathbb{E}\Big[\hat{D}(\hat{D} - \mathbb{E}[\hat{D} \mid X]) \mid X\Big]\right] \\
&= \mathbb{E}\left[LATE(X)\Big[\mathbb{E}[\hat{D}^2 \mid X] - \mathbb{E}[\hat{D} \mid X]^2\Big]\right] \\
&= \mathbb{E}\left[LATE(X)\operatorname{var}(\hat{D} \mid X)\right]
\end{aligned}$$

In the top application of LIE, the first term is killed because all terms are in $\sigma(X, Z)$, so we may condition on X and see that the parenthetical term becomes zero.

Now we consider the denom

$$\begin{aligned}
\operatorname{var}(\tilde{D}) &= \mathbb{E}[\tilde{D}^2] && (\mathbb{E}[\tilde{D} = 0) \\
&= \mathbb{E}[(\hat{D} - \mathbb{E}[\hat{D} \mid X])^2] \\
&= \mathbb{E}\Big[\hat{D}^2 - 2\hat{D}\mathbb{E}[\hat{D} \mid X] + \mathbb{E}[\hat{D} \mid X]^2\Big] \\
&= \mathbb{E}\left[\mathbb{E}\Big[\hat{D}^2 - 2\hat{D}\mathbb{E}[\hat{D} \mid X] + \mathbb{E}[\hat{D} \mid X]^2 \mid X\Big]\right] && \text{(LIE)} \\
&= \mathbb{E}\left[\mathbb{E}\Big[\hat{D}^2 \mid X\Big] - \mathbb{E}\Big[\hat{D} \mid X\Big]^2\right] && (\mathbb{E}[\hat{D} \mid X] \in \sigma(X)) \\
&= \mathbb{E}[\operatorname{var}(\hat{D} \mid X)]
\end{aligned}$$

All that is left to show is that $\hat{D} = p(X, Z)$, which is trivial since D is binary

$$\begin{aligned}
\hat{D} &= \mathbb{E}[D \mid Z, X] \\
&= \mathbb{P}[D = 1 \mid Z, X] \\
&= p(X, Z)
\end{aligned}$$

(d) We will do this piecemeal.

$$\begin{aligned}
\mathbb{E}\Big[\frac{G(1-Z)D}{\mathbb{P}[Z=0 \mid X]}\Big] &= \mathbb{E}[\mathbb{E}\Big[\frac{G(1-Z)D}{\mathbb{P}[Z=0 \mid X]} \mid X\Big]] \\
&= \mathbb{E}[\mathbb{E}\Big[\frac{G(1-Z)D}{\mathbb{P}[Z=0 \mid X]} \mid X, Z=0\Big] \cdot \mathbb{P}[Z=0 \mid X]\Big] && (Z \text{ binary}) \\
&= \mathbb{E}[\mathbb{E}[G(1-Z)D \mid X, Z=0]] \\
&= \mathbb{E}[\mathbb{E}[GD_0 \mid X, Z=0]] && (D \text{ binary}) \\
&= \mathbb{E}[\mathbb{E}[G \mid X, Z=0, D_0=1] \cdot \mathbb{P}[D_0=1 \mid X, Z=0]] && (D \text{ binary}) \\
&= \mathbb{E}[\mathbb{E}[G \mid X, Z=0, D_0=1] \cdot \mathbb{P}[D_0=1 \mid X]] && (\text{Exogeneity}) \\
&= \mathbb{E}[\mathbb{E}[G \mid X, Z=0, D_0=D_1=1] \cdot \mathbb{P}[D_0=D_1=1 \mid X]] && (\text{Monotonicity}) \\
&= \mathbb{E}[\mathbb{E}[G \mid X, D_0=D_1=1] \cdot \mathbb{P}[D_0=D_1=1 \mid X]] \\
& && (\text{Exogeneity (relevant for } Y))
\end{aligned}$$

Similar steps lead us to

$$\mathbb{E}\Big[\frac{GZ(1-D)}{\mathbb{P}[Z=1 \mid X]}\Big] = \mathbb{E}[\mathbb{E}[G \mid X, D_0=D_1=0] \cdot \mathbb{P}[D_0=D_1=0 \mid X]]$$

Now we find

$$\begin{aligned}
\mathbb{E}[\kappa G] &= \mathbb{E}[\mathbb{E}[G \mid X, D_0=0, D_1=1] \cdot \mathbb{P}[D_0=0, D_1=1 \mid X]] \\
&\quad + \mathbb{E}[\mathbb{E}[G \mid X, D_0=1, D_1=0] \cdot \mathbb{P}[D_0=1, D_1=0 \mid X]] \\
& && (\text{Law of Tot. Pr., above calcs}) \\
&= \mathbb{E}[\mathbb{E}[G \mid X, D_0=0, D_1=1] \cdot \mathbb{P}[D_0=0, D_1=1 \mid X]] && (\text{Exogeneity}) \\
&= \int_X \mathbb{E}[G \mid X, D_0=0, D_1=1] \cdot \mathbb{P}[D_0=0, D_1=1 \mid X=x]\mathbb{P}[X=x]\mathrm{d}x \\
&= \int_X \mathbb{E}[G \mid X, D_0=0, D_1=1] \cdot \mathbb{P}[X=x \mid D_0=0, D_1=1]\mathbb{P}[D_0=0, D_1=1]\mathrm{d}x \\
& && (\text{Bayes}) \\
&= \mathbb{P}[T=c] \int_X \mathbb{E}[G \mid X, D_0=0, D_1=1] \cdot \mathbb{P}[X=x \mid D_0=0, D_1=1]\mathrm{d}x \\
&= \mathbb{P}[T=c] \cdot \mathbb{E}[G \mid T=c]
\end{aligned}$$

Dividing the first and final expressions by $\mathbb{P}[T=c]$ gives what we want.

Problem 3

a)

Exogeneity: Conditional on high school grades and year the lotteries are random. Within lottery category and year the instrument is therefore a random process independent of any other (pre-determined) variables. It therefore satisfies exogeneity.

Exclusion: This means that the only way that the lottery should affect wages is through attending medical school. This could fail due to other behavioural responses to the lottery. For example, if those who do not win the lottery choose to reapply the following year in the hope of getting accepted this could cause them to lose experience. In the absence of a lottery they may have begun work immediately or entered into a non-medical degree program. This could affect their earnings, violating exclusion.

Monotonicity: This rules out defiers: those who would go to medical school if they do not win the lotter but would reject the offer to attend if they won the lottery. This seems like a plausible assumption.

b)

We estimate the following first stage:

Table 1: First Stage Regression

	Dependent variable:
	d
z	0.520***
	(0.024)
Constant	0.410***
	(0.023)
Observations	1,476
R^2	0.326
Adjusted R^2	0.326
Residual Std. Error	0.346 (df = 1474)
F Statistic	712.956*** (df = 1; 1474)
Note:	*p<0.1; **p<0.05; ***p<0.01

Some people suggest $F > 10$ as a rule of thumb. There is no real reason for this but $712.956 \gg 10$. The coefficient is highly significant. The instrument appears relevant.

c)

Using IV with Z as an instrument for D produces the following estimates.

Table 2: Instrument Variables Estimates

	Dependent variable:
	lnw
d	0.187***
Constant	3.011***
	(0.023)
Observations	1,476
R^2	0.010
Adjusted R^2	0.010
Residual Std. Error	0.466 (df = 1474)
Note:	*p<0.1; **p<0.05; ***p<0.01

Given that randomisation occurs within category year this coefficient is not easily interpretable in terms of causal effects. Given that the dependent variable is in logs its interpretation would be in percentage change for a unit change in the covariate.

d)

We never know exactly who the compliers are and there will be sampling variation. However we can produce consistent estimates under the assumption of random assignment of Z. Following notation from lecture we have that:

$$p_a = \mathbb{P}(D_1 = D_0 = 1) = \mathbb{P}(D = 1|Z = 0)$$
$$p_n = \mathbb{P}(D_1 = D_0 = 0) = \mathbb{P}(D = 0|Z = 1)$$
$$p_c = \mathbb{P}(D_1 = 1, D_0 = 0) = 1 - p_a - p_c$$

Using the sample relative frequencies to estimate these we find that: $\hat{p}_c = 0.5203$ Which gives an estimated number of compliers of 768.

The share of women in the population is 0.6240.

To find the share of compliers who are women, letting F denote a dummy for female, we note that:

$$\begin{aligned}
\mathbb{P}(F = 1|D_1 > D_0) &= \frac{\mathbb{P}(D_1 > D_0|F = 1)\mathbb{P}(F = 1)}{\mathbb{P}(D_1 > D_0)} && \text{By Bayes's Rule} \\
&= \frac{\mathbb{E}[D_1 - D_0|F = 1]\mathbb{P}(F = 1)}{\mathbb{E}[D_1 - D_0]} && \text{Using Monotonicity} \\
&= \frac{\Big(\mathbb{E}[D|Z = 1, F = 1] - \mathbb{E}[D|Z = 0, F = 1]\Big)\mathbb{P}(F = 1)}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]} && \text{Since } Z \text{ is randomly assigned}
\end{aligned}$$

The first term in the numerator is a first stage regression for women. The denominator is a first stage on the whole population. These results hold by applying the law of total expectation. For example:

$$\begin{aligned}
\frac{Cov(D,Z)}{Var(Z)} &= \frac{\mathbb{E}[DZ]-\mathbb{E}[D]\mathbb{E}[Z]}{\mathbb{E}[Z](1-\mathbb{E}[Z])} \\
&= \frac{\mathbb{E}[D|Z=1]\mathbb{E}[Z]-\Big(\mathbb{E}[D|Z=1]\mathbb{E}[Z]+\mathbb{E}[D|Z=0](1-\mathbb{E}[Z])\Big)\mathbb{E}[Z]}{\mathbb{E}[Z](1-\mathbb{E}[Z])} \\
&= \mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]
\end{aligned}$$

Computing the given regressions we find that the share of compliers that are women is 0.6047. This seems close to the share in the population. One could do a statistical test if so inclined.

e)

In general an IV estimate is a LATE estimate. Even neglecting the issue that randomiation occurs at the category-by-year level the in general we will have LATE≠ATT. This is because in general:

$$LATE = \mathbb{E}[Y_1 - Y - 0|D_1 > D_0] \neq \mathbb{E}[Y_1 - Y_0|D_1 = 1] = ATT$$

The two estimates coincide when the only people to receive treatment are compliers so that $\{D_1 > D_0\} = \{D = 1\}$. However above we estimated $\hat{p}_a > 0$ so that here the two will be different.

f)

We can obtain the mean of Y_1 for compliers using an IV regression of YD on D with Z as instrument since:

$$\mathbb{E}[Y_1|D_1 > D_0] = \frac{\mathbb{E}[YD|Z=1]-\mathbb{E}[YD|Z=0]}{\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]}$$

This holds because:

$$\begin{aligned}
\mathbb{E}[YD|Z=1]-\mathbb{E}[YD|Z=0] &= \mathbb{E}[Y_1-Y_0|D_1>D_0]\mathbb{P}(D_1=1,D_0=0) && \text{Law of total expectation} \\
&= \mathbb{E}[Y_1-Y_0|D_1>D_0]\mathbb{E}[D_1-D_0] && \text{Monotonicity} \\
&= \mathbb{E}[Y_1-Y_0|D_1>D_0](\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]) && \text{Exogeneity}
\end{aligned}$$

Similarly we can get the mean of Y_0 for compliers with an IV regression of $Y(1-D)$ on $(1-D)$ with Z as instrument since similarly to the above:

$$E[Y_1|D_1 > D_0] = \frac{\mathbb{E}[Y(1-D)|Z=1]-\mathbb{E}[Y(1-D)|Z=0]}{\mathbb{E}[1-D|Z=1]-\mathbb{E}[1-D|Z=0]}$$

We estimate the mean of Y_1 for compliers as 3.2642 and the mean of Y_0 for compliers as 3.0770

As in lecture let $f_{zd}(y) = f(y|Z=z, D=d)$ then and let g_{td} be the counterfactual distribution for and individual of type $t \in \{a,c,n\}$ receiving treatment $d \in \{0,1\}$. Note that only g_{n0} and g_{a1} are well defined concepts: always takers always take, never takers never take. We therefore abbreviate these by g_n, g_c respectively. As in class we have the counterfactual distributions for compliers:

$$g_{c0}(y) = f_{00}(y)\frac{p_c + p_n}{p_c} - f_{10}(y)\frac{p_n}{p_c}$$

Which we plot:

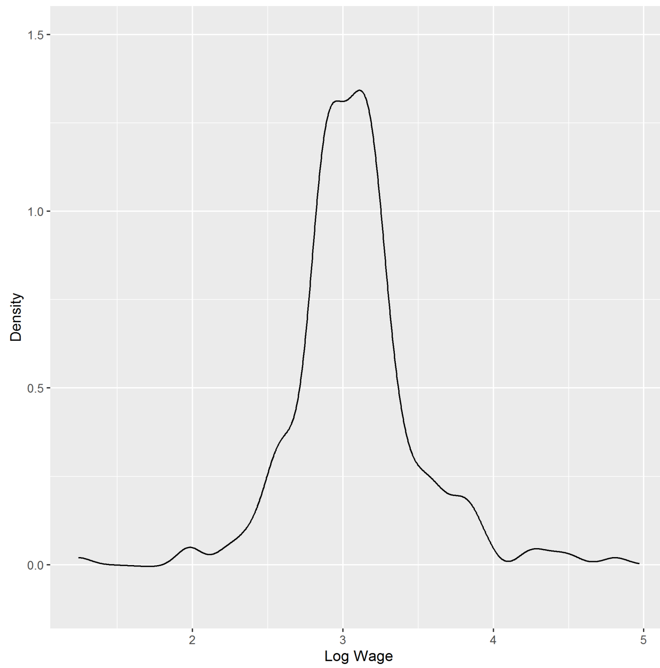

Figure 1: Compliers Y_0 Density

And:

$$g_{c1}(y) = f_{11}(y)\frac{p_c + p_a}{p_c} - f_{10}(y)\frac{p_a}{p_c}$$

Which again we can plot:

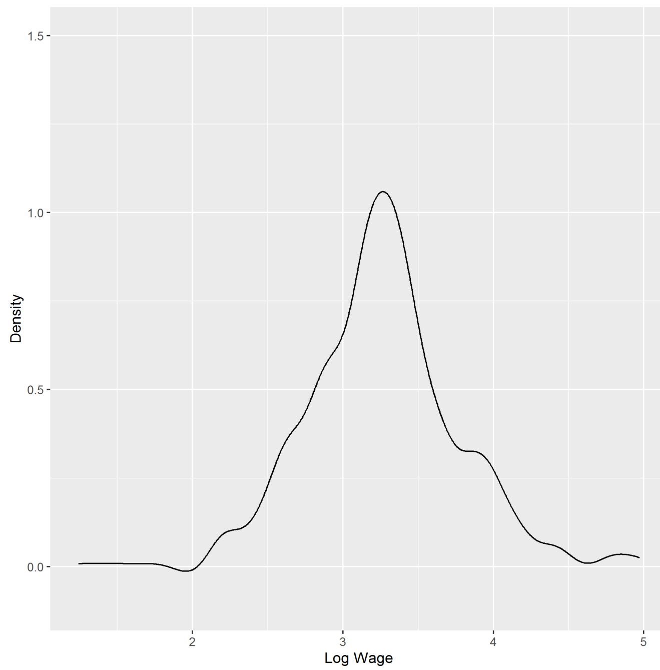

Figure 2: Compliers Y_1 Density

g)

As argued above only g_{n0} and g_{a1} make sense. We plot these below:

For never takers:

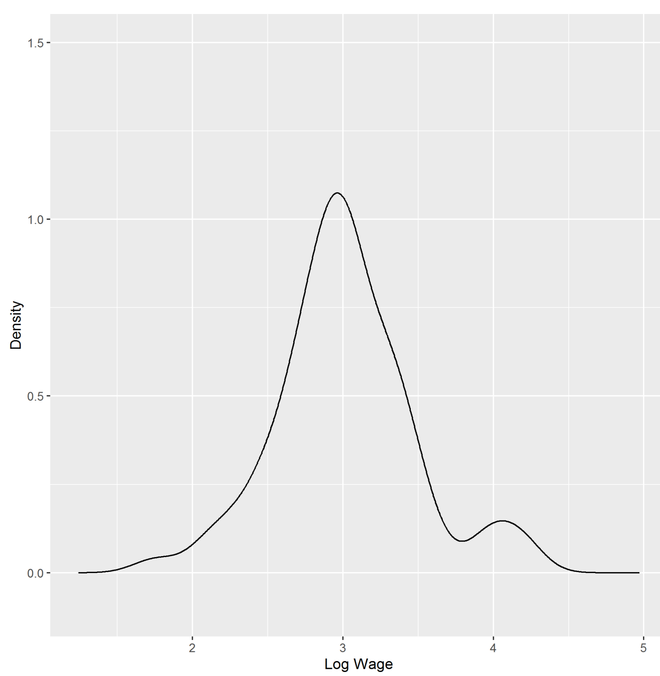

Figure 3: Never Takers Y_0 Density

For always takers:

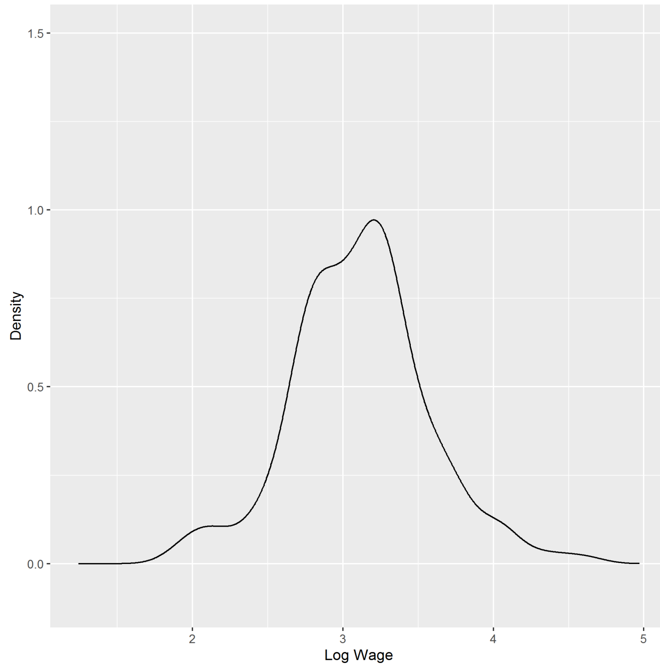

Figure 4: Always Takers Y_1 Density

h)

We estimate each category which gives us the following:

Table 3: LATE Estimates

	1988	1989
3	-0.804	1.564
4	0.120	0.012
5	0.159	0.519
6	0.210	0.145

We aggregate in two ways: using population shares and using complier shares. These produce estimates of 0.2128 and 0.2062 respectively. Running TSLS with a first stage with lottery-category interactions and then a second stage controlling for these gives an estimate of 0.1915. These are all pretty close to each other which makes you wonder whether Angrist & Pischke are always as wrong as all that in practice.

Problem 4

$$ldrugexp = \alpha + \gamma \cdot hiempunion + X\beta + U$$

Where

- *idrugexp*: log expenditure on prescribed medical drugs
- *hi_empunion*: equal to 1 if the individual has supplemental health insurance, 0 otherwise.
- X : age, gender, log of household income (*linc*), number of children (*totchr*), a dummy for being black or hispanic (*blhisp*).

a)

We may be worried about the omitted variable bias, and more precisely that there exists a variable driving both the ownership of supplemental health insurance, and medical drugs expenditures. For example, a poor underlying health condition might lead an individual to buy a supplemental health insurance, knowing they will have high medical drug expenditures. In this case, we would have $E(hiempunion, U) \neq 0$.

b)

Let's see whether *multlc*, a dummy for whether the firm at which the individual is employed is a large operator with multiple locations, could be a good instrument by looking at each of the required assumptions.

1. Random Assignment: $Y_{d,z}, D_z \perp Z|X, \forall d, z$. This assumption would be violated if people working at big companies had different medical expenditures even after controlling for whether or not they have additional health insurance. This condition is likely to be violated because it is implausible that working at a big firm is as good as random after controlling for the covariates. As an example, people working in big cities are more likely to work in big firms, and may also have an urban lifestyle that leads them to spend more on prescribed medical drugs (before they have easier access to a doctor) than people leaving in the countryside, even after controlling for age, income etc.

2. Exclusion Restriction: $Y_{d,1}|X = Y_{d,0}|X = Y_d|X$. This assumption would be violated if there were self-selection into the decision to work at a big firm on the basis of (expected) medical expenditures (but not the availability of supplemental medical insurance), after controlling for covariates. Higher wages (and therefore higher share of income spent on health) in big firms could be one of the drivers, but we control for household income which can be argued to be a (imperfect) proxy for wage. However, note that because where you work, insurance take up and medical expenditures can be seen as a joint decision, the exclusion restriction is likely not to hold.

3. First-Stage / Relevance: $E[D_1 - D_0|X] \neq 0$: it seems reasonable to believe that an individual employed at a big firm is more likely to have supplemental health insurance (as argued above). Indeed, my understanding of the US healthcare system suggests that big firms subsidize (supplemental) health insurance more than small ones. Looking at the third regression table confirms that the first stage is not 0 : the coefficient in front of *multlc* is $0.1488(0.1086, 0.1890)$, for a t-stat of 7.26 and a F-statistic of 120.25.

4. Monotonicity: $P(D_1 \geq D_0|X) = 1$, or vice-versa. This implies that no individual would buy supplemental insurance if working at a small firm, but would not if working at a small firm. As we have argued above, this is not likely to happen because it seems to be easier / cheaper to get supplemental insurance at big firms. Note that if one could show that some smaller firms offer more advantageous / cheaper supplemental health insurance, the argument could be violated.

5. Overlap: $P(Z = 1|X) \in (0, 1)$. This condition ensures that for each value (or group) of X, there are both individuals receiving the instrument ($Z = 1$) and individuals who don't ($Z = 0$). Considering the low number of people employed at big firms in the sample (626 out of 10,089), it could very well be that for each combination of covariates, we don't observe both $Z = 0$ and $Z = 1$. Therefore, this condition could be violated.

Do you think *multlc* is a weak instrument? Among all the conditions discussed above, the one relevant for weak instruments is the first-stage. As argued above, the coefficient for *multl* on the first stage is statistically significant and associated with a F-statistic above 100. Therefore, even if $Cov(multlc, hiempunion) = 0.014051$ is very low, **the instrument is likely not to be weak**.

c)

To obtain the ILS representation of the IV-estimator using *multlc* as an instrument we:

1. Write the equation of interest and the 1st stage in a system of simultaneous equations[5]:

$$\begin{aligned} ldrugexp &= \alpha_1 + \gamma \cdot hiempunion + X\beta_1 + U_1 \\ hiempunion &= \alpha_2 + \sigma \cdot multl + X\beta_2 + U_2 \end{aligned}$$

2. Solve for *ldrugexp* in terms of *multl*

$$\begin{aligned} ldrugexp &= \alpha_1 + \gamma \cdot (\alpha_2 + \sigma \cdot multl + X\beta_2 + U_2) + X\beta_1 + U_1 \\ &= \underbrace{(\alpha_1 + \gamma\alpha_2)}_{\alpha^*} + \underbrace{\gamma\sigma}_{\gamma^*} multl + \underbrace{(\beta_1 + \gamma\beta_2)}_{\beta^*} X + \underbrace{\gamma U_2 + U_1}_{U^*} \end{aligned}$$

3. Look at the regressions of *ldrugexp* on *multl*, and of *hiempunion* on *multlt* and their estimated coefficients from the output tables

$$\begin{aligned} ldrugexp &= 6.0009 - 0.2002194 \cdot multl + \beta^* X + U^* \\ hiempunion &= 0.9017 + 0.1487593 \cdot multl + \beta_2 X + U_2 \end{aligned}$$

4. Recover $\gamma = \frac{\gamma^*}{\sigma} = \frac{-0.2002194}{0.1487593} = -1.34593$.

The estimated $\gamma = -1.34593$ implies that having supplemental health insurance reduces medical drug expenditures by $e^{-1.34593} - 1 = -73.96\%$ among compliers (those who get supplemental insurance because they work at a big firm and who would not if they worked at a small firm).

[5] Do you know IV was first developed to address issues of simultaneous equation systems? How fascinating. See Stock and Trebbi (2003) `https://pubs.aeaweb.org/doi/pdfplus/10.1257/089533003769204416`

d)

Assuming $\beta = 0$, the IV-estimator using the moments (covariances) is equal to

$$\begin{aligned}\gamma &= \frac{Cov(multl, ldrugexp)}{Cov(multl, hiempunion)} \\ &= \frac{-0.016529}{0.014051} \\ &= -1.1763575546\end{aligned}$$

The estimated $\gamma = -1.1763575546$ implies that having supplemental health insurance reduces medical drug expenditures by $e^{-1.1763575546} - 1 = -69.156\%$ among compliers (those who get supplemental insurance because they work at a big firm and who would not if they worked at a small firm).

e)

The share of females in the complier group is the share of females who don't choose extra insurance if their firm is not present in multiple locations, and who choose extra insurance if their firm is present in multiple locations.
We compute the share of female in the complier group by rulling out defiers using the monotonicity assumption.
Using Bayes Rule we know that

$$\begin{aligned}Pr(f|c) &= \frac{Pr(f,c)}{Pr(c)} \\ &= \frac{Pr(c|f)Pr(f)}{Pr(c)}\end{aligned}$$

where f refers to female, c refers to compliers. Below, a refers to always-takers, n refers to never-takers.

Let's first compute $Pr(c|f)$, i.e. the probability of being a complier given being female:

	Insured	Multiple 0	Multiple 1
Female:	0	3,721	125
	1	1,792	184

$$\begin{aligned}p_{c|f} &= 1 - p_{a|f} - p_{n|f} \\ &= 1 - Pr(D=1|Z=0,f) - Pr(D=0|Z=1,f) \\ &= 1 - \frac{Pr(D=1,Z=0,f)}{Pr(Z=0|f)} - \frac{Pr(D=0,Z=1|f)}{Pr(Z=1|f)} \\ &= 1 - \frac{\frac{1792}{5822}}{\frac{1792+3721}{5822}} - \frac{\frac{125}{5822}}{\frac{125+184}{5822}} \\ &= 1 - \frac{0.307798}{0.946925} - \frac{0.02147}{0.0530745} \\ &= 1 - 0.325 - 0.4045 \\ &= 0.2705\end{aligned}$$

Therefore $Pr(c|f) = 0.2705$, which corresponds to $0.2705 \cdot 5,822 = 1,575$ people.
Let's do the same with the whole population to compute $Pr(c)$ Adding men and women:

Insured	Multiple 0	Multiple 1
0	3,721 + 2,267	125 + 120
1	1,792 + 1,683	184 + 197

i.e.

Insured	Multiple 0	Multiple 1
0	5,988	245
1	3,475	381

$$\begin{aligned}
p_c &= 1 - p_a - p_n \\
&= 1 - Pr(D=1|Z=0) - Pr(D=0|Z=1) \\
&= 1 - \frac{Pr(D=1, Z=0)}{Pr(Z=0)} - \frac{Pr(D=0, Z=1)}{Pr(Z=1)} \\
&= 1 - \frac{\frac{3475}{10089}}{\frac{3475+5988}{10089}} - \frac{\frac{245}{10089}}{\frac{245+381}{10089}} \\
&= 1 - \frac{0.34443}{0.93795} - \frac{0.02428}{0.06205} \\
&= 1 - 0.367 - 0.3913 \\
&= 0.2417
\end{aligned}$$

Finally, we know that $Pr(f) = \frac{f}{f+m} = \frac{5822}{10089} = 0.577$.

Therefore we have

$$\begin{aligned}
Pr(f|c) &= \frac{Pr(f,c)}{Pr(c)} \\
&= \frac{Pr(c|f)Pr(f)}{Pr(c)} \\
&= \frac{0.2705 \cdot 0.577}{0.2417} \\
&= 0.6457
\end{aligned}$$

This affects the interpretation of the estimates because it means that the estimates on the compliers are driven by female observations at 64.57% (versus 57.7% in the population). Women are over-represented among the compliers .

f)

From before we showed that :
Among women

- The share of always takers : $p_{a|f} = \frac{Pr(D=1,Z=0,f)}{Pr(Z=0|f)} = 0.325$
- The share of never takers : $p_{n|f} = \frac{Pr(D=0,Z=1|f)}{Pr(Z=1|f)} = 0.4045$
- The share of compliers : $p_{c|f} = 1 - p_{a|f} - p_{n|f} = 0.2705$

We follow the same approach and show that, **among men**

- The share of always takers : $p_{a|m} = \frac{Pr(D=1,Z=0,m)}{Pr(Z=0|m)} = 0.4261$
- The share of never takers : $p_{n|m} = \frac{Pr(D=0,Z=1|m)}{Pr(Z=1|m)} = 0.3785$
- The share of compliers : $p_{c|m} = 1 - p_{a|m} - p_{n|m} = 1 - 0.4261 - 0.3785 = 0.1954$

Overall we have shown that

- The share of always takers $p(a) = 0.367$
- The share of never takers $p(n) = 0.3913$
- The share of compliers $p(c) = 0.2417$

Therefore, we have that

- **The share of women in the complier group is**

$$\begin{aligned} p(f|c) &= \frac{p(f,c)}{p(c)} \\ &= \frac{p(c|f)p(f)}{p(c)} \\ &= \frac{0.2705 \cdot 0.577}{0.2417} \\ &= 0.6457 \end{aligned}$$

- **The share of women in the always-taker group is**

$$\begin{aligned} p(f|a) &= \frac{p(f,a)}{p(a)} \\ &= \frac{p(a|f)p(f)}{p(a)} \\ &= \frac{0.325 \cdot 0.577}{0.367} \\ &= 0.51097 \end{aligned}$$

- **The share of women in the never-taker group is**

$$\begin{aligned} p(f|n) &= \frac{p(f,n)}{p(n)} \\ &= \frac{p(n|f)p(f)}{p(n)} \\ &= \frac{0.4045 \cdot 0.577}{0.3913} \\ &= 0.5965 \end{aligned}$$

g)

From the last table we have

$$E(Y_0|n) = E(Y_0|D=0, Z=1) = 6.029153$$
$$E(Y_1|a) = E(Y_1|D=1, Z=0) = 6.558737$$

To compute $E(Y_0|c)$ and $E(Y_1|c)$ we need to do a little more work (as always).
From the lecture notes we have that

$$E(Y_1|c) = \frac{E(YD|Z=1) - E(YD|Z=0)}{E(D|Z=1) - E(D|Z=0)}$$
$$E(Y_0|c) = \frac{E(Y(1-D)|Z=1) - E(Y(1-D)|Z=0)}{E(1-D|Z=1) - E(1-D|Z=0)}$$

Looking at $E(Y_1|c)$: Rearranging to adjust for the fact that we don't have a regression of Y D on Z:

$$\begin{aligned}
\hat{E}(Y_1|c) &= \frac{E(YD|Z=1) - E(YD|Z=0)}{E(D|Z=1) - E(D|Z=0)} \\
&= \frac{[E(YD|D=1, Z=1)P(D=1|Z=1) + E(YD|D=0, Z=1)P(D=0|Z=1)]}{P(D=1|Z=1) - P(D=1|Z=0)} \\
&- \frac{[E(YD|D=1, Z=0)P(D=1|Z=0) + E(YD|D=0, Z=1)P(D=0|Z=1)]}{P(D=1|Z=1) - P(D=1|Z=0)} \\
&= \frac{E(Y|D=1, Z=1)P(D=1|Z=1) - E(Y|D=1, Z=0)P(D=1|Z=0)}{P(D=1|Z=1) - P(D=1|Z=0)} \\
&= \frac{E(Y|D=1, Z=1)\frac{381}{381+245} - E(Y|D=1, Z=0)\frac{3475}{3475+5988}}{\frac{381}{381+245} - \frac{3475}{3475+5988}} \\
&= \frac{6.3345 \cdot \frac{381}{381+245} - 6.5587 \cdot \frac{3475}{3475+5988}}{\frac{381}{381+245} - \frac{3475}{3475+5988}} \\
&= \frac{6.3345 \cdot 0.6086261981 - 6.5587 \cdot 0.3671528218}{0.6086261981 - 0.3671528218} \\
&= \frac{1.4472974395}{0.2414733763} = 5.9936
\end{aligned}$$

Similarly for $E(Y_0|c)$:

$$\begin{aligned}
\hat{E}(Y_0|c) &= \frac{E(Y(1-D)|Z=1) - E(Y(1-D)|Z=0)}{E(1-D|Z=1) - E(1-D|Z=0)} \\
&= \frac{E(Y|D=0, Z=1)P(D=0|Z=1) - E(Y|D=0, Z=0)P(D=0|Z=0)}{P(D=0|Z=1) - P(D=0|Z=0)} \\
&= \frac{E(Y|D=0, Z=1)\frac{245}{245+381} - E(Y|D=0, Z=0)\frac{5988}{5988+3475}}{\frac{245}{245+381} - \frac{5988}{5988+3475}} \\
&= \frac{6.029153 \cdot \frac{245}{245+381} - 6.464303 \cdot \frac{5988}{5988+3475}}{\frac{245}{245+381} - \frac{5988}{5988+3475}} \\
&= \frac{6.029153 \cdot 0.3913738019 - 6.464303 \cdot 0.6327803022}{0.3913738019 - 0.6327803022} \\
&= \frac{-1.730831074}{-0.2414065003} = 7.16978
\end{aligned}$$

For external validity, finding $E(Y_0|c) \neq E(Y_0|c)$ and $E(Y_1|c) \neq E(Y_1|a)$ means that the estimated effect of the program is only relevant for complier-type people. It might not be relevant to generalize it to the whole population. Not only are $E(Y_0|c) \neq E(Y_0|c)$ and $E(Y_1|c) \neq E(Y_1|a)$, but we even have $E(Y_1 - Y_0|c) < 0$ while $E(Y_1|a) - E(Y_0|n) > 0$...