

# Sheet 5

outstanding

(a)  $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + u$ , satisfying OLS.

- g) Consider running a short regression of  $Y$  on  $X_1, \dots, \tilde{X}_{k-1}$ .  
 By FWL, the coefficient on  $X_1$ ,  $\gamma_1 = \frac{\text{Cov}(\tilde{X}_1, Y)}{\text{Var}(\tilde{X}_1)}$ , where  
 $\tilde{X}_1$  is the residual in a regression of  $X_1$  on  $X_2, \dots, X_{k-1}$ .

Using FWL just got me confused here, but

seems like it

Should be relevant?

From our long model above, we can substitute in to the numerator

$$\gamma_1 = \frac{\text{Cov}(\tilde{X}_1, \beta_0 + \sum_{l=1}^k \beta_l X_l + u)}{\text{Var}(\tilde{X}_1)} = \frac{\text{Cov}(\tilde{X}_1, \beta_0) + \text{Cov}(\tilde{X}_1, \beta_1 X_1) + \text{Cov}(\tilde{X}_1, \beta_2 X_2) + \dots + \text{Cov}(\tilde{X}_1, u)}{\text{Var}(\tilde{X}_1)}$$

*Yep. see* ~~so is a constant so its cov with  $\tilde{X}_1$  is 0. Therefore~~

*in fact for a derivation via FWL*

Run a population regression of  $X_k$  onto  $X_1, \dots, X_{k-1}$ , obtaining OLS coefficients  $\pi_0, \pi_1, \dots, \pi_{k-1}$ :  $X_k = \pi_0 + \pi_1 X_1 + \dots + \pi_{k-1} X_{k-1} + v$

We can rewrite the structural equation as

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{k-1} X_{k-1} + \beta_k (\pi_0 + \pi_1 X_1 + \dots + \pi_{k-1} X_{k-1} + v) + u$$

$$= (\beta_0 + \beta_k \pi_0) + (\beta_1 + \beta_k \pi_1) X_1 + \dots + (\beta_{k-1} + \beta_k \pi_{k-1}) X_{k-1} + (\beta_k v + u)$$

$\checkmark$       !!      !!      !!      !!      !!

$\beta_0$        $\beta_1$        $\beta_{k-1}$        $v$

To show that a short regression of  $Y$  on  $X_1, \dots, X_{k-1}$  recovers the above  $\gamma_1$ , we need to demonstrate that  $v$  is orthogonal to each  $X_1, \dots, X_{k-1}$ , and mean zero.

*Very nice* | By construction,  $E[u] = E[v] = 0$ , so  $E[\epsilon] = 0$ . And also by construction  $v$  is orthogonal to each  $X_1, \dots, X_{k-1}$  (as is  $u$ ), so  $\epsilon$  satisfies this too, by Cov properties. So, instead,

$$\gamma_1 = \beta_1 + \beta_k \pi_1$$

and  $\beta_k > 0$

- b)  $\gamma_1 = \beta_1 + \beta_k \pi_1$ . If  $\text{cov}(X_1, X_k) > 0$ , then - intuitively - some of the positive effect that  $X_k$  has on  $Y$  ~~then~~ is misattributed to  $X_1$ , given that  $X_1$  moves with  $X_k$  but none of the other regressors do. So by omitting  $X_k$  from the short regression, we overestimate the true effect of  $X_1$  on  $Y$ .

More formally, note that  $\pi_1 = \frac{\text{cov}(X_k, \tilde{x}_1)}{\text{var}(\tilde{x}_1)}$  by FWL,

where  $\tilde{x}_1$  is the residual in a regression of  $X_1$  on  $X_2, \dots, X_{k-1}$ .  
i.e.  $X_1 = \gamma_0 + \gamma_2 X_2 + \dots + \gamma_{k-1} X_{k-1} + \tilde{x}_1$

And  $\text{cov}(X_k, \tilde{x}_1) = \text{cov}(X_k, X_1) - \text{cov}(X_k, \gamma_0) - \text{cov}(X_k, \gamma_2 X_2) - \dots$  ✓  
 $= \text{cov}(X_k, X_1) - 0 - 0 \xrightarrow[\text{by assumption } \gamma_0 \text{ is constant}]{} \dots > 0$

Since  $\text{var}(\tilde{x})$  is weakly positive,  $\pi_1 > 0$ , so  $\beta_k \pi_1 > 0$ , so  $\gamma_1 > \beta_1$ .  
(and nonzero to not have  
the regression degenerate)

attrition

- a) This doesn't affect internal validity, because you can just consider the trial to have started out without those male athletes, assuming they were evenly split between the treatment and control groups, which (However, if e.g. all the athletes ~~tried nearby to each other~~ somehow ended randomisation should've achieved).

- b) This is contamination and does affect internal validity, because some people in the control group received treatment and thus the difference in group means can't be interpreted as the treatment effect.  
(You'd get an underestimate, probably.)

- c) This is non-compliance and, again, does affect internal validity, because some people in the treatment group didn't actually get treated. (However, if you're trying to merely estimate the FTT effect, then this doesn't matter, ~~much~~, ~~except~~ because you will still recover the average treatment effect on those you intended to treat)

Great

- d) Similar to (b), this creates contamination. Framed differently, access to treatment is no longer  $\perp$  of ~~moderated~~<sup>pre-treatment</sup> characteristics (e.g. wealth, "grit") and so observed group difference  $\neq$  ATE (even if you knew that these individuals managed to access treatment). Also, maybe the econ majors ended up with slower internet speeds, which is similar to non-compliance in (c).

- e) Not really a major threat. If you didn't know which rooms were affected then you'd probably underestimate the true  $\beta$  (when there aren't storm disruptions because some treated people didn't get a "full" treatment), but you could just omit the storm-affected people from your analysis because the randomness of it means that disruption  $\perp$  pre-treatment characteristics.

5. In the sample regression of  $Y$  on  $D$ , we have the model

$$Y = \hat{\beta}_0 + \hat{\beta}_1 D + \hat{u} \quad \text{where } E[\hat{u}] = E[\hat{u}|D] = 0,$$

and  $\hat{\beta}_1 = \frac{\text{cov}(Y, D)}{\text{var}(D)}$

$$= \hat{E}[(Y - \bar{Y})(D - \bar{D})] \div \hat{E}[(D - \bar{D})^2]$$

$$= (\hat{E}[YD] - \hat{E}[Y]\hat{E}[D]) \div \hat{E}[(D - \bar{D})^2]$$

$$= \left( \frac{1}{n} \sum_{\substack{i=1 \\ D_i=1}}^n Y_i - \left( \frac{1}{n} \sum_{i=1}^n Y_i \right) \left( \frac{1}{n} \sum_{i=1}^n D_i \right) \right) \div \underbrace{\frac{1}{n} \sum_{i=1}^n (D_i - \frac{n_i}{n})^2}_{\text{This is just the sample variance of } \text{Bern}\left(\frac{n_i}{n}\right), \text{ which}}$$

$$= \frac{1}{n} \sum_{\substack{i=1 \\ D_i=1}}^n Y_i - \frac{1}{n^2}$$

is  $\frac{n_i/n \cdot ((1-n_i)/n)}{n} = \frac{n_i n_{-i}}{n^2}$ ?

6a) Regression (3) explores whether we can predict treatment assignment based on a variety of pre-treatment characteristics. The purpose here is to test for balances and check if randomisation was successful. We can infer from the F-Statistic that these coefficients are not significant even at the 10% level (ca. for  $F_{6, 38} = 1.77 > 1.20$ ) and so they don't predict treatment status. i.e., based on observed variables at least, randomisation was successful.

(not inconsistent with  $D \perp\!\!\!\perp X$ )

b) What's of interest to the policymaker here is how income transfers affect food consumption, because they'd potentially affect food consumption, i.e.,  $\frac{\partial F}{\partial I} \cdot \frac{I}{F}$ . You can't make a  $\Delta$ -sized change to income, but you can make a small  $\delta$ -sized change to it and observe how  $F$  changes, which is what the study does through estimating the OLS coefficient.

**Yes** [Using household income as your independent variable, <sup>would</sup> undermines this because it isn't randomly assigned and very plausibly differs due to other unobserved characteristics that affect food consumption, e.g. occupation. It also means you get a worse approximation to a  $\Delta$ -sized change in  $I$ .]

c) Yes, you can give this ~~as~~ a causal interpretation - it's reasonable to think that the residual is uncorrelated with income transfer (since we had random assignment), i.e. OLR holds. The 95% CI is  $[0.65 \pm 1.96 \times 0.123]$ , since  $P(|N(0,1)| > 1.96) = 0.05$ , i.e.  $[0.42, 0.90]$ . This are the values  $b$  for which we could not reject  $H_0: \beta = b$  against  $H_1: \beta \neq b$  at the 5% level, based on this sample.

d) Introducing the additional regressors is not necessary for consistency of estimating  $\beta$  (we have that already as  $y_{nc}$ ), but rather improving precision in our estimate of  $\beta$ . Our new OLS coefficient is

(I don't find  
this intuitive  
story totally  
satisfying, even  
though the maths  
makes sense)

Slightly smaller, but not in a significant way given the SE on the original OLS estimate! However, the SE on that point estimate reduces ~~very~~ substantially (by about 30%). Intuitively this is because when we allow the model to include relevant pre-treatment covariates of  $F$ , it can achieve much better fit than with  $I$  alone, and these smaller residuals allow us to be less uncertain about the true causal effect of  $I$ .

childhood

e) ✓ Height has been included as a proxy for health, so it doesn't make sense to interpret its OLS coefficient causally. There's no sensible story for why it would appear in the causal model. Childhood Health, however, very plausibly does affect food consumption, e.g. people spend relatively ~~more~~ <sup>less</sup> on food ~~because~~ <sup>because</sup> better. ↗ who grew up malnourished ↗ because their bodies are accustomed to it.

f) Note that for all control group individuals,  $I=0$  but  $\log(I)$  is undefined, so this won't really work. Also, we care about elasticity wrt total income, not transfer income. ↗  
As discussed in (b), you could instead calculate the elasticity at ~~the~~ a certain point by substituting in to  $\frac{\partial F}{\partial I} \cdot \frac{I}{F}$  with the partial derivative  $\frac{\partial F}{\partial I} = \beta$ , and then evaluate at the average household total income  $\bar{H}$  and food consumption  $\bar{F}$ , for all households in the study.

3. (Alternatively, because  $I$  was continuous, we could just ~~the~~ regress  $\log(I)$  onto  $\log(F)$  for households who received  $I > 0$ . But this is still elasticity about the wrong thing, I think, and also reduces our sample size. In addition, maybe there's a fixed effect from receiving any transfer, which perhaps we miss accounting for like this, although I feel like even with all the data we wouldn't be capturing it, given the model used?)

7a) If we carry out the regression  $Y = \gamma_0 + \gamma_1 D + \epsilon$ , we can be reasonably happy that  $\gamma_1$  recovers the causal effect of treatment on earnings, because it means  $D$  was assigned randomly. So, we should have  $D \perp\!\!\!\perp u$ , assuming randomisation was successful, and therefore omitted variables folded into  $u$  - like family background - shouldn't cause any bias.

It's worth noting that this study is over an extremely long time period. We might be concerned about the precision on our estimates, due to high attrition rates, extremely large variability in  $u$ . An approach which adds additional covariates and/or incorporates short-run outcome data might be preferable.

either how we define, or

- b) Let's assume for the sake of argument that  $X$  is a good proxy for educational attainment. (It very well might not be - e.g. students might spend many years in school because they fail to graduate; years in education isn't perfectly associated with attainment)

~~It seems rather unduly attack~~

~~By FWL, in our regression  $Y = \beta_0 + \beta_1 D + \beta_2 X + u$ ,~~

~~is FWL not relevant here?~~

$$\beta_1 = \frac{\text{cov}(\tilde{D}, Y)}{\text{var}(\tilde{D})} \quad \text{where } \tilde{D} \text{ is the residual in a regression of } D \text{ on } X$$

If kindergarten class size really does affect earnings, part of this causal chain might be mediated via years in education, i.e.  $D \rightarrow X \rightarrow Y$ .

Blah \*Bear was expert that in this case, it doesn't make sense to talk about the causal effect of  $D$  "holding  $X$  constant", since  $X$  is not actually constant, given a change in  $D$ !

Substituting in our auxiliary regression,

$$Y = \beta_0 + \beta_1 D + \beta_2 (\delta_0 + \delta_1 D + v) + u$$

$$= (\beta_0 + \beta_2 \delta_0) + (\beta_1 + \beta_2 \delta_1) D + (\beta_2 v + u)$$

$$=: \gamma_0 + \gamma_1 D + \epsilon \quad \text{as in (a).}$$

In fact, this OLS won't

yield consistent causal estimates at all:

Moreover,  $\text{cov}(u, X) = \text{cov}(u, \delta_0 + \delta_1 D + v) = \text{cov}(u, v) \neq 0$  because, e.g., ability is an unobserved determinant of both attainment and earnings.

So then our regression  $Y = \beta_0 + \beta_1 D + \beta_2 X + u$  will not yield consistent estimators of the causal effects, since OR fails: we're using a bad control of attainment which is endogenous. (However, maybe the OLS coefficient  $\beta_1$  would be consistent for the direct causal effect of  $D$  on  $Y$ ? Although I'm not sure about how to interpret the "holding  $X$  constant", I guess we can think of it as just "Supposing  $X$  didn't vary with  $D$ "?)

*possibly  
be it  
 $u \perp\!\!\!\perp v$*

In principle, it feels like you could control on an intermediate outcome (i.e. something "endogenous") while still having  $u \perp\!\!\!\perp v$ . (maybe if you put in enough other proxies / covariates.) Is that not true / not true in practice? This reminds me of some work I did before on the "surrogate index", but I don't totally remember it.

5. From our sample regression  $Y = \hat{\beta}_0 + \hat{\beta}_1 D + \hat{u}$ , we have

$$\mathbb{E}[\hat{u}] = \mathbb{E}[\hat{u}|D] = 0, \text{ and } \hat{\beta}_1 = \frac{\hat{\text{cov}}(Y, D)}{\hat{\text{var}}(D)}$$

$$\begin{aligned}\hat{\text{cov}}(Y, D) &:= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})(D_i - \bar{D}) = \frac{1}{n} \sum Y_i D_i + \bar{Y} \bar{D} - \frac{\bar{Y}}{n} \sum D_i - \frac{\bar{D}}{n} \sum Y_i \\ &= \frac{1}{n} \sum_{\{i : D_i = 1\}} Y_i + \cancel{\frac{n_1}{n} \bar{Y}} - \cancel{\frac{n_1}{n} \bar{Y}} - \cancel{\frac{n_1}{n} \bar{Y}} \\ &= \frac{1}{n} \left( \sum_{\{i : D_i = 1\}} Y_i - n_1 \bar{Y} \right)\end{aligned}$$

Hmm, I'm just going to leave this here. The lecture notes say it is "tedious but easy algebra"; I'm not sure how easy it is but I agree otherwise...