Sheet 3

1a) The coefficient is relatively small, at $|\hat{\beta}| \approx 0.1$: holding ~~the other~~ meduc ~~and~~ feduc constant, it'd need to decrease by 10.6 to increase expected educ by a year. ✓

It seems reasonable that we predict more years of education for men in smaller families: it depends where the data are from, but e.g. if you have to pay for school in a country, more siblings likely means less money for one individual's education.

But the small effect size might make it hard to reliably conclude that this is the true population sign of the coefficient.

b) Holding sibs and feduc constant, on average an increase in 1 yr of mother's educ is associated with a 0.131 yr increase in son's education in this sample.

Good

This probably isn't a reliable causal estimate, because it's likely that in the underlying causal model, immigration status is a confounder: women from certain countries are likely to have had much less time in education, and children's national background might directly affects their yrs in education ~~also~~ e.g. due to cultural reasons. ✓

c) $\widehat{educ}(A) - \widehat{educ}(B) = 0.131 \times -4 + 0.210 \times -4$
$= -1.364$ yrs ✓

d) As described in (b), factors like nationality + religion may affect both educ and m/f educ. Also in this category is parents' family wealth (esp. as it's likely highly correlated with the man's family wealth growing up), and ~~and~~ child's ability. Sibling order is correlated with sibs and probably affects educ. Perhaps child's ~~[illegible]~~ health also affects educ, and would likely be correlated with the others too. [note that a lot of this is tied up to ✓

also perhaps corr w/ sibs as poorer families tend to be larger (there are various other "random" things ⊥ of the other variables which might affect educ, e.g. ~~[illegible] exit school~~ being called for military service)

whatever gets OR to become plausible.

wealth / income. If you just added those, could you happily ignore the rest? e.g. plausibly after controlling for £, child health ⊥ parental educ ?]

2a) Older workers have more experience / pay progression so very likely earn more: on average, in the sample, being 1yr older is associated with roughly 50¢ /hr higher wage.
holding gender and degree constant

It doesn't really make sense to treat this as a causal effect, but it's a good proxy for what is the likely causal lever: years of labour market experience. If you took someone and let them age in a cave for 10yrs that wouldn't increase their ~~predicted~~ expected wage...

(or I guess more directly: productivity)

b) There's a large negative effect on predicted wages from being female: it's associated with almost $4/h lower wages.

holding age and degree constant

This is slightly surprising: although the gender pay gap is a significant phenomenon, normally it only emerges after women return to the workforce after having children, aged around 35 or older, but this sample doesn't include such people.
[just had a look at ONS data, in the UK currently it's ~1% for this age range but has previously been reversed].
Given that this is all full-time work, i.e. not about annual pay etc, one explanation could be that women are more likely to be in lower-paid sectors than men. (or that they're simply paid less for equal work, though that's not legal.)

c) No, the coeff almost certainly doesn't give a causal estimate. People with degrees are generally more academically able, healthier, from higher-income / education backgrounds and these all affect expected earnings directly. You'd want to either introduce these as additional observed proxied variables, or (perhaps better) use some kind of quasiexperimental setup like FDP with people who normally don't get accepted for a degree programme.

3 a) Note that there is perfect multicollinearity between {sleep, study, work, leisure}, because they must sum to 168. So if you try to vary one without changing the others, you're in effect making a prediction for someone who has more/fewer than 24 hours in a day, which clearly doesn't make sense. The "ceteris paribus" framing for interpreting β_1 isn't possible, so it's not appropriately interpreted as a causal effect.

b) No, the OLS regression will not be able to produce $\hat{\beta}_i$ for ~~these~~ this model because of the problem above. The FWL theorem says $\hat{\beta}_i = \frac{Cov(Y, \tilde{x}_i)}{Var(\tilde{x}_i)}$, but when you attempt to find the coefficient for the final variable, the residual from the auxiliary regression will be 0 (always) and so the denominator here is undefined. Put differently, there are infinitely many combinations of coefficients which would minimise sample MSE, so the optimisation problem is degenerate.

c) Just drop e.g. sleep as an explanatory variable, so there's no longer the collinearity problem. Then, you can interpret each remaining $\hat{\beta}$ as giving the expected effect of swapping one hour of sleep ~~for~~ the respective activity, i.e. relative to our baseline.

5 a) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$, $E[u \mid X_1, X_2] = 0$

a) A population linear regression of Y on X_1 and X_2 yields the following model:

$$Y = \rho_0 + \rho_1 X_1 + \rho_2 X_2 + e, \text{ where by construction}$$
$$E[e] = E[X_1 e] = E[X_2 e] = 0.$$

The goal is to show that the ρ_i and β_i coincide, by demonstrating that the causal coefficients satisfy the normal equations uniquely characterising the ~~no~~ OLS coefficients from FOCs, namely ~~namely~~ $\mathbb{E}[\underbrace{Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2}_{\overset{!!}{=} e}] = \mathbb{E}[(\cdot)X_1] = \mathbb{E}[(\cdot)X_2] = 0$.

oh, or use LIE $\mathbb{E}[\mathbb{E}[u|X_1, X_2]]$ ✓ Yep

- Since $\mathbb{E}[u|X_1, X_2] = 0$, $\sum_{\varepsilon} [\varepsilon \cdot P(u=\varepsilon | X_1 = x_1, X_2 = x_2)] = 0$ for all x_1, x_2. But then $\sum_{x_2} \sum_{x_1} \sum_{\varepsilon} \varepsilon \cdot P(u=\varepsilon | X = x_1, X_2 = x_2) \cdot P(X_1 = x_1, X_2 = x_2) = \underline{0 = \mathbb{E}[u]}$ by the law of total probability.

- $$\mathrm{cov}(X_1, u) \equiv \underset{\|}{\mathbb{E}[X_1 u]} - \mathbb{E}[X]\underset{\overset{\|}{0}}{\underline{\mathbb{E}[u]}}$$

$$\cancel{= \sum_{x_1, \varepsilon} [\varepsilon \cdot X_1 \cdot P(X_1 = x_1 \cap u = \varepsilon)]}$$

you could add in conditioning on X_2 to be more explicit

$$= \mathbb{E}[\mathbb{E}[X_1 u | X_1]] \quad \text{by LIE}$$

$$= \mathbb{E}[X_1 \underbrace{\mathbb{E}[u|X_1]}_{0}] \quad \text{by conditioning}$$

$$= \underline{0} \quad \text{by assumption.}$$ ✓

b) No, not unless $\beta_2 = 0$, or [illegible] $X_1 \perp\!\!\!\perp X_2$. If you regress Y on X_1 alone, we'll get $\beta_1' = \frac{\mathrm{Cov}(X_1, Y)}{\mathrm{Var}(X_1)}$, where according to the causal model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$.

Nice!

So $\mathrm{Cov}(X_1, Y) = \beta_1 \mathrm{Var}(X_1) + \beta_2 \mathrm{Cov}(X_1, X_2) + \underset{\overset{\|}{0}}{\underline{\mathrm{Cov}(X_1, u)}}$

And thus $\beta_1' = \beta_1 \iff \beta_2 = 0$ or $\mathrm{Cov}(X_1, X_2) = 0$

Intuitively, if X_2 is relevant to Y and correlated with X_1, then when you omit it from the regression, the OLS ~~estimate~~ coefficient on X_1 will change to incorporate how changes in X_1 can predict changes in X_2, which in turn causally affect Y.

7. $Y_i^* = aY_i + c \qquad X_i^* = bX_i$

a) ~~By~~ We know that $\hat{\beta}_1 = \dfrac{\hat{Cov}(X, Y)}{\hat{Var}(X_1)}$ and similarly $\hat{\beta}_1^* = \dfrac{\hat{Cov}(X_1^*, Y^*)}{\hat{Var}(X_1^*)}$

$$\therefore \hat{\beta}_1^* = \frac{\hat{Cov}(bX_1, aY + c)}{\hat{Var}(bX_1)} = \frac{ab\,\hat{Cov}(X_1, Y)}{b^2\,\hat{Var}(X_1)} = \underline{\frac{a}{b}\hat{\beta}_1}$$

And for $i = 0$, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$; $\hat{\beta}_0^* = \bar{Y}^* - \hat{\beta}_1^*\bar{X}^*$

$$\therefore \hat{\beta}_0^* = (a\bar{Y} + c) - \left(\frac{a}{b}\hat{\beta}_1\right)\cdot(b\bar{X}) = a(\bar{Y} - \hat{\beta}_1\bar{X}) + c = \underline{a\hat{\beta}_0 + c}$$

b) $SER := \sqrt{\dfrac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2}$ where in this case $k = 1$.

It's an estimate of the standard deviation of the residuals u_i.

Because $\hat{u}^* := Y^* - (\hat{\beta}_0^* + \hat{\beta}_1^* X^*)$

$$= aY + c - (a\hat{\beta}_0 + c + \tfrac{a}{b}\hat{\beta}_1 bX)$$

$$= a(Y - (\hat{\beta}_0 + \hat{\beta}_1 X)) =: a\hat{u},$$

$\hat{Var}(\hat{u}^*) = a^2\,\hat{var}(\hat{u})$ and so the SER will increase by a factor of $|a|$ (as k is unchanged)

i.e. the translation c and rescaling of X by b doesn't affect SER but rescaling of Y does, since it

c) $R^2 := \dfrac{ESS}{TSS}$ where $ESS := \sum_i(\hat{Y}_i - \bar{Y})^2$; $TSS := \sum_i(Y_i - \bar{Y})^2$

But considering R^{2*}, note that the ~~both~~ ESS ~~and~~ $\div$ TSS ratio will be unchanged

$$ESS^* = \sum_i(\hat{Y}_i^* - \bar{Y}^*)^2 = \sum_i\left(a\hat{\beta}_0 + c + \tfrac{a}{b}\hat{\beta}_1 bX - (a\bar{Y} + c)\right)^2$$

$$= a^2\sum_i(\hat{Y}_i - \bar{Y})^2$$

Similarly $TSS^* = a^2 TSS$, and so the scaling cancels out.

Good quest. See in class.

I'm not sure what sort of model this is. Causal? but why -i? Sample regression wouldn't make sense.

~~So the R^2~~ isn't sensitive to affine transformations of either Y or X_{12}. ✓ Great.

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. (I) which satisfies OR

oh wait, this was talking about a sample, not population?

First perform the auxiliary regression of X_1 onto X_2 in the sample

$$X_{1i} = \hat{\pi}_0 + \hat{\pi}_2 X_{2i} + \hat{\tilde{X}}_{1i}$$

where by construction, $\sum_{i=1}^{n} [\tilde{X}_{1i}] = \sum_{i=1}^{n} [\tilde{X}_{1i} X_{2i}] = 0$ (*)

Now, substituting into (I),

but I'm confused by the mix of β (without hats) and $\hat{\pi}$ here

$$Y_i = \beta_0 + \beta_1 (\hat{\pi}_0 + \hat{\pi}_2 X_{2i} + \tilde{X}_{1i}) + \beta_2 X_{2i} + u_i$$

$$= \underbrace{(\beta_0 + \beta_1 \hat{\pi}_0)} + \underbrace{(\beta_1 \hat{\pi}_2 + \beta_2)} X_{2i} + (\beta_1 \tilde{X}_1 + u_i)$$

$$=: \gamma_0 + \beta_1 \tilde{X}_{1i} + \gamma_2 X_{2i} + u_i \quad \text{(II)}$$

~~Regressing Y onto X_{2i},~~
~~when you find the OLS coefficient $\hat{\beta}_1$, you'll have a model~~

There's something I don't get about this question... e.g. where is it that we "do" the Y on $\tilde{X}_1$ regression? My (II) isn't a regression model, right - it's the auxiliary regression substituted into the economic model (I).

A regression of Y_i onto $\tilde{X}_1, X_2$ would produce a model $Y_i = \hat{\beta}_0 + \hat{\beta}_1$

From (II), $\hat{Cov}(Y, \hat{\tilde{X}}_1) = \hat{Cov}(\hat{\tilde{X}}_1, \gamma_0) + \hat{Cov}(\hat{\tilde{X}}_1, \beta_1 \hat{\tilde{X}}_1) + \hat{Cov}(\hat{\tilde{X}}_1, \gamma_2 X_2 + u_i)$

$$= 0 + \beta_1 \hat{Var}(\hat{\tilde{X}}_1) + 0$$

since $\hat{Cov}(\hat{\tilde{X}}_1, X_2) = 0$ by (*)
and $\hat{Cov}(\hat{\tilde{X}}_1, u) = \hat{Cov}(X_1 - \hat{\pi}_0 - \hat{\pi}_2 X_2, u)$
$= \hat{Cov}(X_1, u) - \hat{Cov}(\hat{\pi}_0, u) - \hat{\pi}_2 \hat{Cov}(X_2, u)$
$= \underset{\text{by (I)}}{0} - \underset{\text{constant}}{0} - \underset{\text{by (I)}}{0}$

$$\therefore \beta_1 = \frac{\hat{cov}(Y, \hat{\tilde{X}}_1)}{\hat{Var}(\hat{\tilde{X}}_1)}$$

(there's a hat missing on β_1, clearly, but where was it meant to appear?)