Sheet 4

2a) The test statistic $t = \frac{\beta_1^{obs} - \beta_1^{H_0}}{se(\beta_1^{obs})}$ where $\beta_1^{H_0} = 0$, β_1^{obs} is the ~~value of~~ variable $\hat{\beta}_1$ in your sample regression, and $se(\beta_1^{obs})$ is an estimate of the standard deviation of $\hat{\beta}_1$, ($\hat{\omega}_{\beta_1}/\sqrt{n}$ with sample size n and asymptotic variance $\hat{\omega}_{\beta_1}$)

I don't have a good intuition about what ω (or $\hat{\omega}$) is, where it comes from, etc. Can we discuss in class?

Under H_0 and by the CLT, $t \sim N(0,1)$ in a large sample. So we reject H_0 iff $t^{obs} > c_\alpha$, where c_α is pinned down s.t. $P(N(0,1) > c_\alpha) = \alpha = 10\%$.

yep.

b) i. The smaller the value of β_1 in the population, the harder it is for the test to detect whether observed data sampled from that population is not consistent with H_0. This is because the t-statistic's numerator will be small, as $\beta_1 \approx \beta_1^{H_0}$.

ii. It will never correctly detect that H_0 should be rejected if $\beta_1 = -1$ (since it would still have Type I errors) because the one-sided test only looks for unexpectedly high values of t, not all large ones incl. -ve.

c) For each true β_1, the power will be slightly less (perhaps something like the Type II error rate doubles?) because now the ~~critical~~ t-values which get rejected are spread out on both tails. However, it will now be able to detect if $\beta_1 = -1$, with symmetric power to $\beta_1 = 1$.

4a) $R^2 := \frac{ESS}{TSS} = 0.134$ ✓ $\bar{R}^2 := 1 - \frac{n-1}{n-k-1} \cdot \frac{SSR}{TSS}$ where $SSR = TSS - ESS$ and $k = 4$

$= 0.132$ ✓

As expected in this large sample with a relatively small number of regressors, $R^2 \approx \bar{R}^2$ ✓

$SER := \left(\frac{1}{n-k-1} SSR\right)^{1/2} = 3.72$ ✓

$H_0: \beta_1 = \beta_2 = \cdots = \beta_q = 0$. $\quad H_1: \beta_l \neq 0$ for some $l \in [1, q]$

To find the F-statistic, we fit a model
$Y = \beta_0^{rs} + u^{rs}$, where in this case OLS would arrive at $\beta_0^{rs} = 61.81$.

In the restricted regression, the SSR will be related to the square of the SD: $= \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$ because for each observation, our model predicts $\bar{Y}$. Specifically, we need to scale the sample variance by $(n-1)$, so $SSR = 3.99^2 \times 6027$.

But this is simply $= TSS$, by definition.

$$\text{So } F := \frac{n-k-1}{q} \cdot \frac{SSR_{un} - SSR_{rs}}{SSR_{un}} = \frac{6018}{9} \cdot \frac{83097 - 95948}{83097}$$

$= 103.4$

Looking at c_α for a 5% test with $F_{9,\infty}$, we have a decision rule to reject H_0 if $F^{obs} > 1.88$, and so we reject H_0; the model isn't useless.

b) $H_0: \beta_8 = 0 \quad H_1: \beta_8 \neq 0 \quad$ where β_8 is the population OLS coeff on teacher exp.

$$t := \frac{\hat{\beta}_8 - \beta_8^{H_0}}{se(\hat{\beta}_8)} \overset{obs}{=} \frac{0.04}{0.02} = 2$$

maybe better to use different notation for the RV. and its realization

By CLT in large samples, under H_0 $t \sim N(0,1)$.
We reject H_0 iff $|t| > c_\alpha$ where $P(|N(0,1)| > c_\alpha) = 5\%$
i.e. $c_\alpha = 1.96$.

So as $2 > 1.96$, reject H_0.

c) The p-value is the probability of observing a sample at least as averse to H_0 as the one we did, supposing H_0 to be true.

In this case, we can find $t^{obs} = 2.2$, and calculate

$$P(|t| > t^{obs}) \approx P(|N(0,2)| > 2.2) = \cancel{2\times} P(N(0,2) \leq 2.2) = \cancel{0.027}$$

$$\underline{0.014}$$

Great.

Alternatively, we can interpret this as the smallest α at which we could've rejected H_0 on the basis of this sample.

d) The 99% CI is the range of values of β_3^{obs} which would be consistent with $H_0: \beta_3 = 0$ at a 1% significance level against $H_1: \beta_3 \neq 0$.

$\phi^{-1}(0.005) = -2.58$, so we want to include a range 2.58 std. devs around the mean, s.t. $P(\beta_3^{obs}$ not in range $| H_0) = 0.01$.

i.e. the 99% CI is $[-0.853, -0.287]$? If so, not quite right since the interval is not random.

Alternatively we can interpret this as an interval which has 99% chance of containing β_3, across all samples. ...meaning this

see in class

e) We're interested in $\left.\dfrac{\partial\, maths\_test / maths\_test}{\partial\, teacher\_exp / teacher\_exp}\right|_{(61.81,\ 13.93)}$.

From our OLS output, we can see that a 1-unit ↑ in exp is associated with a 0.04 ↑ in score, on average in the sample. So the elasticity is $0.04 \times \dfrac{13.93}{61.81} = 0.0090$.

The 90% CI stretches $\pm \phi^{-1}(0.05) \times se(\varepsilon)$ around 0.0090, where $se(\varepsilon) \cancel{=} se(\hat{\beta}_8)$ holding the sample means fixed,
$= 0.02$

i.e. $[-0.024, 0.042]$

$se(\varepsilon) = se\left(\dfrac{13.93}{61.81}\hat{\beta}\right) = \ldots$

~~f) ... We find $z^{obs} = -9.86$ for $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ which is significant at the~~

f) We <u>could</u> approach this by performing two t-tests for β_1 and β_2, looking at the p-values, and/or seeing whether the $(1-\alpha)$% CIs cross 0. But it's better (?) to use an F-test.

So, one thing abt F tests is that they only really tell you that <u>something</u> is not insignificant, right? Yep.

$H_0: \beta_1 = \beta_2 = 0 \qquad H_1: \beta_1 \neq 0$ or $\beta_2 \neq 0$

We'd ideally want to re-run the restricted regression and collect the statistics, using R.

Here, we can immediately say that β_1 is significant at the 1% level ($t^{obs} = -9.56$), so being Black ~~has~~ a ~~negative~~ associated with lower scores, holding constant the other variables and relative to being White. For β_2, $t^{obs} = 1.53$, so ~~only~~ not significant at 5% level, or even 10%.

g) No, obviously that's silly, these are correlation coeffs and OR is unlikely to hold because the unobserved variable av-school-income probably causes ↑ FSM and ↓ scores, but isn't proxied for by any of the other regression variables, so the coeff on FSM doesn't have a causal interp.

(Also, only 47% of children have FSMs, so even if the interp ~~for~~ we causal, removing FSMs wouldn't make average test scores go up by 1.79 as only a minority of students would see a change in circumstances).

5a) $C_i + T_i + R_i = 1$ as they're MECE. $(\forall i)$

b) i. The average change in w_i associated with 1 yr more experience, holding constant location.

ii. —— " ———————————— living in the city relative to {living in the rurally ~~or~~/ in a town} respectively. holding experience constant.

For i, these will be equal upto sampling variation between (3) and (4); for ii the coefficients are just about different things.

c) $T_i = 1 - C_i - R_i$ as noted in a)

So we can rewrite (3) as

$$W_i = \beta_0 + \beta_X X_i + \beta_C C_i + \beta_T(1 - C_i - R_i) + u_i$$

$$= \underbrace{(\beta_0 + \beta_T)} + \beta_X X_i + \underbrace{(\beta_C - \beta_T)} C_i - \beta_T R_i + u_i$$

$$(4) \quad W_i = \gamma_0 + \gamma_X X_i + \gamma_C C_i + \gamma_R R_i + v_i$$

i.e. $\gamma_R = -\beta_T$, so while in (3) β_T tells us the ~~effect~~ change on W (holding X_i constant) associated with being in the town vs a baseline of countryside, in (4) γ_R gives us this same number just moving in the other direction.

d) $H_0: \beta_C = \beta_T = 0 \quad H_1: \beta_C \neq 0$ or $\beta_T \neq 0$.

As in previous Qs, we can now calculate the F-statistic by fitting a restricted regression model and comparing the outputs to an unrestricted model.

Then, we use our decision rule to reject H_0 iff $F^{obs} > c_\alpha$ where $P(F_{?,?} > c_\alpha) = \alpha$, ~~as dtz the experienced conditioned~~

2?

e) You could produce a new model with interaction terms:

$$W_i = \beta_0 + \beta_X X_i + \beta_C C_i + \beta_T T_i + \beta_{CX}(C_i \cdot X_i) + \beta_{TX}(T_i \cdot X_i) + u_i$$

(Do we need both? Also how do you interpret interaction terms for continuous variables?)

Cross-partial derivative.

Then produce OLS estimates of the interaction term coefficients, and perform a hypothesis test ~~that~~ against a null that these coefficients are in fact both 0. (Again, you'd use an F-test.)

Oh also, if you add new interaction terms could OR fail to hold even if it did before? (~~intuitively~~ (My guess would be yes?) But what's the reasoning?

... in your regression? Doesn't affect OR...

I remember asking before and possibly just forgot the answer — Sorry — but is there any significance to the 'i' subscripts appearing in models vs. not?

No. Not particularly, but we can discuss the point again in class. Pls bring it up.

6. $Y_i = A_i L_i^{\alpha} K_i^{\beta} \cdot \exp(\varepsilon_i)$

a) Note that $\log Y_i = \log\left[A L_i^{\alpha} K_i^{\beta} \cdot \exp(\varepsilon_i)\right]$

$$= \log A + \alpha \log L_i + \beta_1 \log K_i + \varepsilon_i$$

So you can run a regression of log Y on log L and log K, recovering α and β as the coefficients, ~~with log A our intercepts and ε_i the residuals.~~

~~was~~ (Because ε_i is ~~mean~~ uncorrelated with L and K, our OLS estimates will be consistent for the true parameters. It doesn't matter that it's not mean zero, except that we can't read off the intercept as log A)

b) The Cobb-Douglas prod. func. is CRS if $\alpha + \beta = 1$. So we can just test whether our sample OLS estimates are consisted with this null hypothesis.

We could perform a t-test on $H_0: \gamma := \hat{\alpha} + \beta = 1$ with $\gamma^{obs} := \alpha^{obs} + \beta^{obs}$ ~~and although we need either need~~ and so $var(\hat{\gamma}^{obs}) = var(\hat{\alpha}) + var(\hat{\beta}) + \underbrace{2\,Cov(\hat{\alpha}, \hat{\beta})}_{\text{unclear how to calculate; probably R can?}}$

Alternatively, perform an F-test where the restricted model is of the form $\log Y_i = \pi_0 + \alpha \log L_i + (1-\alpha)\log K_i + u_i$ i.e. $Y^* = \pi_0 + \pi_1 X_1 + u$ with $Y^* = (\log Y_i - \log K_i)$, $\pi_1 = \alpha$ and $X_1 = (\log L_i - \log K_i)$

Not sure what you mean by this.

c) ~~We~~ The approach in (a) would lead us to omit the variation in A_i (since our intercept term would merely be $\bar{A}$), and this variation would be rolled into ε_i. But then OR is not satisfied in the causal model and OLS won't consistently recover the relevant coefficients. The direction of bias would

is that true? about $\bar{A}$

depend on the precise correlation structure.

(Can we be more specific / is there a version of OVB formula for multiple variables, not just X_1 and omit X_2?)

→ Yes. You can use FWL.

d). With panel data, you could add dummy variables for each country (leaving out one to baseline) that are there to capture the inter-country variation in A. Then OR would be satisfied and you can recover α and β.

(However, with only $n=2$ for each dummy coefficient, presumably there will be large SEs around each one?).

This wouldn't have worked with the cross-sectional data, because there we only have one observation for each country so the normal equations end up degenerate.
(As a next-best step, you could've grouped countries by region / economic development and added dummies for that, plausibly proxying the variation in A_i).

[Apparently this is called a "fixed effects" model]

7. [1] a) is true by inspection. Nowhere do the parameters β_i interact or appear except ~~in~~ summations ~~other with~~ as coefficients on functions of X_1 and X_2.

b) i. $\mathbb{E}[u] = \mathbb{E}[uX_1] = \mathbb{E}[uX_1^2] = \mathbb{E}[uX_2] = \mathbb{E}[uX_1X_2] = 0$

ii. $\mathbb{E}[u \mid X_1, X_2] = 0$ (it's not enough to say $\mathbb{E}[u|X_1] = \mathbb{E}[u|X_2] = 0$)

[2] a) again is clear. e.g. we could write

$$Y = [\beta_0, \beta_1, \beta_2, \beta_3] \cdot [1, \log X_1, X_2, X_2 \log X_1]^T + u$$

b) i. $\mathbb{E}[u] = \mathbb{E}[u \log X_1] = \mathbb{E}[u X_2] = \mathbb{E}[u X_2 \log X_1] = 0$

ii. $\mathbb{E}[u \mid X_1, X_2] = 0 \iff \mathbb{E}[u \mid \log X_1, \log X_2] = 0$