Econometrics
Supervision 3

Samuel Lee

Question 1

(a)

Given that this is the estimated relationship,

$$\log(\text{FOOD}) = 4.7377 + 0.3506\log(\text{PDI}) - 0.5086\log(\text{PRICE})$$

then the effects of PDI and PRICE on FOOD are as such

$$\frac{\partial\log(\text{FOOD})}{\partial\text{PDI}} = \frac{1}{\text{FOOD}}\cdot\frac{\partial\text{FOOD}}{\partial\text{PDI}} = \frac{0.3506}{\text{PDI}}$$
$$\frac{\partial\text{FOOD}}{\partial\text{PDI}}\frac{\text{PDI}}{\text{FOOD}} = 0.3506$$

The left-hand side is the elasticity of FOOD with respect to PDI, so a percentage increase in PDI is associated with an increase of approximately 0.3506% in FOOD. Following the same reasoning, a percentage increase in PRICE is associated with a decrease of approximately 0.5086% in FOOD.

(b)

With $H_0 : \beta_{\text{PRICE}} = 0$, $H_1 : \beta_{\text{PRICE}} \neq 0$, the t-statistic for this test is $\frac{\hat{\beta}_{\text{PRICE}}}{\text{s.e.}(\hat{\beta}_{\text{PRICE}})} = -\frac{0.5086}{0.1010} = -5.03564$. With $n - k - 1 = 25 - 2 - 1 = 22$ degrees of freedom, the rejection rule for a two-tailed test is $|t| > 2.074$ at the 5% significance level. The difference from 0 is significant at the 0.1% significance level for that matter. Therefore H_0 is rejected.

(c)

Assuming log(INCOME) actually refers to log(PDI), $H_0 : \beta_{\text{PDI}} = 1$, $H_1 : \beta_{\text{PDI}} \neq 1$. Just from inspection it is apparent that the coefficient in this test is further from H_0 than in (b), and furthermore the standard error is lower. We should expect H_0 to be rejected in that case. Just to verify this, the t-statistic for this test is $\frac{\hat{\beta_{\text{PDI}}} - \beta_{\text{PDI},H_0}}{\text{s.e.}(\hat{\beta}_{\text{PDI}})} = \frac{0.3506-1}{0.0899} = -7.2236$. Unsurprisingly, the absolute value of this t-statistic is larger than that in (b), and H_0 is rejected.

(d)

To clarify, SST is the total sum of squares, SSR is the residual sum of squares, and SSE is the explained sum of squares. The SST is equal to

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n}(y_i^2-2y_i\bar{y}+\bar{y}^2) \\
&= \sum_{i=1}^{n}[(\hat{y}_i+\hat{u}_i)^2-2(\hat{y}_i+\hat{u}_i)\bar{y}+\bar{y}^2] \\
&= \sum_{i=1}^{n}(\hat{y}_i^2+2\hat{y}_i\hat{u}_i+\hat{u}_i^2-2\hat{y}_i\bar{y}-2\hat{u}_i\bar{y}+\bar{y}^2) \\
&= \sum_{i=1}^{n}[(\hat{y}_i-\bar{y})^2+\hat{u}_i^2+2\hat{y}_i\hat{u}_i-2\hat{u}_i\bar{y}] \\
&= \text{SSE}+\text{SSR}+2\sum_{i=1}^{n}(\hat{y}_i\hat{u}_i-\hat{u}_i\bar{y}) \\
&= \text{SSE}+\text{SSR}+2\sum_{i=1}^{n}\hat{y}_i\hat{u}_i && \left(\text{since } \sum_i \hat{u}_i=0 \text{ from the FOC for OLS}\right) \\
&= \text{SSE}+\text{SSR}+2\sum_{i=1}^{n}\sum_{j=0}^{k}\hat{\beta}_j x_{ji}\hat{u}_i && (\text{where } x_k \text{ are the regressors, } x_0=1) \\
&= \text{SSE}+\text{SSR} && \left(\text{since } \sum_i x_{ji}\hat{u}_i=0\ \forall j \text{ from the FOC}\right)
\end{aligned}$$

With that out of the way, the SSE is just $\text{SST}-\text{SSR}=0.52876-0.0046276=0.5241324$, and the R^2 is $\frac{\text{SSE}}{\text{SST}}=\frac{0.5241324}{0.52876}=0.9912$.

(e)

For this a test for the exclusion restrictions for PDI and PRICE must be done. The restricted model in this case has no regressors, and thus $\hat{y}_i=\bar{y}\ \forall i$. R_{r}^2 (the R^2 of the restricted model) in this case is just 0 since SSE$=\sum_i(\hat{y}_i-\bar{y})=\sum_i(\bar{y}-\bar{y})=0$.

The F-statistic for the test on exclusion restrictions is

$$F=\frac{(R_{\text{ur}}^2-R_{\text{r}}^2)/q}{(1-R_{\text{ur}}^2)/(n-k-1)}$$

where q is the number of restrictions, R_{ur}^2 is the R^2 on the unrestricted model, and R_{r}^2 is the R^2 on the restricted model as mentioned before. We know that $R_{\text{r}}^2=0$, so with $H_0:\beta_{\text{PDI}},\beta_{\text{PRICE}}=0, H_1:\beta_{\text{PDI}}$ and/or $\beta_{\text{PRICE}}\neq 0$, the F-statistic just becomes

$$F=\frac{0.9912/2}{(1-0.9912)/22}=1239$$

which is of course rejected at the 5% level.

Question 2

(a)

The results from regressing IQ on parents' education (*feduc* and *meduc*) are as follows

	(1)
	IQ
feduc	1.017***
	(5.40)
meduc	1.062***
	(4.82)
_cons	80.21***
	(38.45)
N	722
F	61.72

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Both coefficients are significantly different from 0. Furthermore, the F-statistic for the test of the exclusion restriction for *feduc* and *meduc* is very large (61.72). Therefore the parents' education levels are both individually and jointly significant, with a positive correlation between IQ and either parent's education in years.

(b)

Adding *educ* to the regression would give the effects of *feduc* and *meduc* on *IQ* conditional on *educ* being held constant. The F-test from excluding *feduc* and *meduc* from this regression would suggest whether the parents' education levels are jointly significant. If there were any channel through which the parents' education levels contribute to IQ other than the daughter's own education, *feduc* and *meduc* should be insignificant.

	(1)	(2)	(3)
	IQ	IQ	IQ
feduc	1.017***	0.363*	
	(5.40)	(2.06)	
meduc	1.062***	0.612**	
	(4.82)	(3.05)	
educ		3.030***	3.547***
		(13.22)	(17.08)
Constant	80.21***	50.38***	53.66***
	(38.45)	(17.19)	(18.67)
R-squared	0.147	0.314	0.288
N	722	722	722

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(1) is the original regression, (2) is the regression with *educ* included and (3) is the restricted model of (2) with *feduc* and *meduc* excluded. For (3) the observations where *feduc* or *meduc* were not available were not considered; the command used was:

```
regress IQ educ if feduc~=.  & meduc~=.
```

From (2) we can see *feduc* and *meduc* are still individually significant although their magnitudes have decreased substantially. Now *feduc* is significant at the 5% level but not the 1% level while *meduc* is now significant at the 1% level but not the 0.1% level.

The F-statistic to test whether *feduc* and *meduc* are jointly significant after including *educ* is

$$F = \frac{(R^2_{(2)} - R^2_{(3)})/2}{(1 - R^2_{(2)})/(722 - 3 - 1)} = \frac{(0.314 - 0.288)/2}{(1 - 0.314)/718} = 13.6064$$

which can be rejected at the 5% significance level.

(c)

The parents' education levels have a statistically significant association with the IQ of their daughters in this sample, controlling for the effect of their education levels on their daughters' education level.

(d)

It is difficult to derive any policy recommendation from 3 simple regressions such as these. For one it is very likely that *feduc* and *meduc* are correlated with the error term. It is likely that parents with more education earn more in general, and this can have effects on the IQ of their daughters. Even if the number of years of schooling is controlled for, children in higher-income households likely benefit from higher quality education, for example through supplementary classes or extra-curricular activities, all of which can affect IQ (or at least the measurement of IQ). Furthermore statistical significance tells us little about economic significance, and in policy-making the important thing is to look at the magnitudes of the purported effects and compare them with other uses of public money.

(e)

The variables $blackeduc = black \times educ$, $blackfeduc = black \times feduc$, and $blackmeduc = black \times meduc$ were generated. The results from the regression of IQ on *educ*, *feduc*, *meduc*, *blackeduc*, *blackfeduc*, and *blackmeduc* are below

	(1)
	IQ
educ	3.060***
	(13.80)
feduc	0.185
	(1.04)
meduc	0.421*
	(2.05)
blackeduc	-1.573***
	(-3.62)
blackfeduc	0.545
	(0.97)
blackmeduc	0.329
	(0.55)
_cons	55.00***
	(19.12)
N	722

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

If we take the partial derivative of the estimated model with respect to *educ*, we get $\frac{\partial IQ}{\partial educ} = \hat{\beta}_{educ} + \hat{\beta}_{blackeduc} \cdot black$. This means the association between *educ* and *IQ* changes depending on whether *black*=0 or *black*=1. The same goes for *feduc* and *meduc*. Thus we are trying to see if the coefficients on *blackeduc*, *blackfeduc*, and *blackmeduc* are significantly different from 0, to determine whether race affects the association between education and IQ. In this case only *blackeduc* has a statistically significant coefficient, which is negative for that matter, meaning the association between education and IQ is weakened if *black*=1.

Question 3

For some $\hat{\theta}$ which is an estimator of θ based on a sample of size n, $\hat{\theta}$ is a consistent estimator of θ if

$$\mathrm{P}(|\hat{\theta} - \theta| > \varepsilon) \to 0 \text{ as } n \to \infty \ \forall \varepsilon$$

or in other words, as the sample size goes to infinity, the probability that the estimated value of θ differs from the true value goes to 0. If we know that $\hat{\theta}$ is an unbiased estimator of θ, and $\mathrm{Var}(\hat{\theta}) \to 0$ as $n \to 0$, then $\hat{\theta}$ must be consistent: $\mathrm{Var}(\hat{\theta}) \to 0$ implies $\hat{\theta}$ is converging on some value, and if $\hat{\theta}$ is unbiased the only value it can converge to is θ.

In this case, the estimator is unbiased (as shown 2 weeks ago) but not consistent. This is because it only uses 3 observations no matter how large the sample size is, and the variance of the estimator does not change however large n gets. If there is some positive variance in the estimator, this means that there is a chance that the estimator will deviate from the true value. Since this is true for all n, $\hat{\beta}_1$ cannot be a consistent estimator.