2024-25 Tripos IIA Paper 3

Supervision 2

Emre Usenmez

Michaelmas 2024

Very grateful to Dr Oleg Kitov and Dr Clive Lawson for the very informative stylized answers to previous iterations of the supervision questions.

FACULTY QUESTIONS

QUESTION 1

In order to understand the determinants of years of schooling among women in the UK, we collected a random sample of 857 individual females aged 28-38 from across the UK and gathered information on a number of background characteristics including their score on an IQ test. The summary statistics are as follows.

Here, *educ* represents years of schooling, *age* is recorded in years, *meduc* is number of years of schooling for the individual's mother, and *IQ* is the score on an IQ test. *black* is a dummy variable identifying whether the woman has African origin. Define *leduc* to be the natural log of years of education.

The OLS regression output with *leduc* as the dependent variable is reported below. For the regression we have $N = 857, R^2 = 0.33$.

(a) Interpret the coefficient 0.005 on IQ.

Answer: Lets first write down the model specification:

$$leduc_i = \beta_0 + \beta_1 IQ_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 black_i + \beta_5 meduc_i + u_i$$

Table 1: Summary statistics for explanatory variables

	Mean	Std Dev	Min	Max
educ	13.58	2.20	9	18
age	32.98	3.09	28	38
black	0.12	0.32	0	1
meduc	10.68	2.85	0	18
IQ	101.80	15.01	50	145

Table 2: OLS regression for 'leduc' as the dependent variable

	Coeff	Std err	p-value
IQ	0.0050	0.0003	0.000
age	0.1202	0.0361	0.001
agesq	-0.0018	0.0005	0.001
black	-0.0314	0.0152	0.039
meduc	0.0126	0.0017	0.000
constant	-0.0569	0.5981	0.924

The OLS estimator of β_1 is the partial effect of IQ holding other factors fixed:

$$\left. \frac{\partial leduc_i}{\partial IQ} \right|_{age, black, meduc} = \beta_1$$

So, holding age, race, and mother's education constant, a one point increase in IQ increases education by 0.5%.

$\overline{}$

(b) How would you test if educational attainment is affected by age in this model?

Answer: First, notice that we have two explanatory variables that depend on age. So we want to test *age* and *agesq* having no effect on educational attainment against either *age* or age^2, or both, having effect on educational attainment. That is,

$$\mathbb{H}_0 : (\beta_2 = 0) \cap (\beta_3 = 0)$$

$$\mathbb{H}_1 : (\beta_2 \neq 0) \cup (\beta_3 \neq 0)$$

To conduct the test, we would first estimate the restricted model

$$\theta_0 + \theta_1 IQ_i + \theta_4 black_i + \theta_5 meduc_i + v_i$$

and use the $F - test$ statistic that follows a $F_{q,n-k-1} = F_{2,857-5-1} = F_{2,851}$ distribution:

$$F = \frac{\frac{(RSS_{res} - RSS_{unr})}{q}}{\frac{RSS_{unr}}{n-k-1}} = \frac{\frac{(RSS_{res} - RSS_{unr})}{2}}{\frac{RSS_{unr}}{851}} \sim F_{q,n-k-1}$$

or,

$$F = \frac{\dfrac{(R^2_{unr} - R^2_{res})}{q}}{\dfrac{1 - R^2_{unr}}{n - k - 1}} = \frac{\dfrac{(0.33 - R^2_{res})}{2}}{\dfrac{1 - 0.33}{851}} \sim F_{q,n-k-1}$$

where, RSS_{unr} and RSS_{res} are the sums of squared residuals from the unrestricted and restricted models, respectively; and R^2_{unr} and R^2_{res} are the coefficients of determination of the unrestricted and restricted models, respectively.

Why F-test?

A joint hypothesis is where two or more restrictions are imposed on the regression coefficients. To test a joint hypothesis, F-statistic is used. It may seem possible to use a t-statistic to test the restrictions one at a time, but this approach would be unreliable.

To see this, suppose that t_1 is the t-statistic for testing the null hypothesis that $\beta_2 = 0$ in this question and t_2 is the t-statistic for $\beta_3 = 0$. Do we then reject the joint null hypothesis $\mathbb{H}_0 : (\beta_2 = 0) \cap (\beta_3 = 0)$ if either t_1 or t_2 exceeds 1.96 in absolute value?

Answering this requires characterizing the joint sampling distribution of t_1 and t_2. $\hat{\beta}_2$ and $\hat{\beta}_3$ have a joint normal distribution due to large sample, so under the joint null hypothesis t_1 and t_2 have a bivariate normal distribution where each t-statistic has a mean of 0 and a variance of 1.

Now consider the special case where t_1 and t_2 are uncorrelated and thus independent in large samples. $\mathbb{H}_0$ is not rejected only if both $|t_1| \le 1.96$ and $|t_2| \le 1.96$.

Because the t_1 and t_2 are independent,

$$\mathbb{P}(|t_1| \le 1.96 \text{ and } |t_2| \le 1.96) = \mathbb{P}(|t_1| \le 1.96) \times \mathbb{P}(|t_2| \le 1.96)$$
$$= 0,95^2 = 0.9025 = 90.25$$

So the probability of rejecting $\mathbb{H}_0$ when it is true is $1 - 0.95^2 = 9.75$

    $\hookrightarrow$ If the regressors are correlated, then the size of this one at a time appraoch, i.e. the probability of rejecting the null when it is true, depends on the magnitude of that correlation.

Since using t-statistics one at a time results in null rejection rates that don't match the desired significance level, another approach is needed.

One approach is called the **Bonferroni method** modifies this t-statistics approach so that it uses critical values that ensure that its size equals its significance level. However, this method can have low power. That is, it can frequently fail to reject the null hypothesis when the alternative is true.

This is why F-test is preferred in testing joint hypothesis. It is more powerful, especially when the regressors are highly correlated. When the joint null hypothesis has two restrictions ($q = 2$) the F-statistic combines the two t-statistics using the formula

$$F = \frac{1}{2} \left(\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}} \right)$$

where $\hat{\rho}_{t_1,t_2}$ is an estimator of the correlation between the two t-statistics.

In the general case of q restrictions, and if the error u_i is homoskedastic, the F-statistic is expressed as

$$F = \frac{\dfrac{(RSS_{res} - RSS_{unr})}{q}}{\dfrac{RSS_{unr}}{n - k - 1}}.$$

This can be thought of as F measuring the relative increase in RSS when moving from unrestricted to restricted model. Also notice that the denominator of F, $\dfrac{RSS_{unr}}{n - k - 1}$ is the unbiased estimator of $Var(u) = \sigma_u^2$.

Another way to think of the question F-statistic addresses is to ask whether relaxing the q restrictions improves the fit of the regression sufficiently that this improvement is unlikely to be the result of random sampling variation if the null is true. This means there is a relationship between F-statistic and R^2. A large F-statistic should be associated with a substantial increase in the regression R^2.

↪ If the error u_i is homoskedastic, the F-statistic can be written in terms of the improvement in the fit of the regression measured either by increase in R^2 or decrease in the residual sum of squares.

This F-statistic and the one that uses RSSs are sometimes referred to as **homoskedasticity-only F-statistic** because it is valid only if the error term is homoskedastic.

This relationship between F-statistic and R^2 is then expressed as

$$F = \frac{\dfrac{(R^2_{unr} - R^2_{res})}{q}}{\dfrac{1 - R^2_{unr}}{n - k - 1}}.$$

Also notice that if we differentiate $leduc_i$ with respect to age_i

$$\frac{\partial leduc_i}{\partial age_i} = \beta_2 + 2\beta_3 age_i$$

then we see that partial effect of age depends on age.

In addition, since $\hat{\beta}_2 = 0.1202 > 0$ and $\hat{\beta}_3 = -0.0018 < 0$, the relationship between natural log of education and age is *concave*.

We can also calculate the age for which the partial effect of age on log education becomes negative:

$$age_i = \frac{-\hat{\beta}_2}{2\hat{\beta}_3} = \frac{-0.1202}{2 \times -0.0018}$$

```
0.1202/(2*0.0018)
```

```
## [1] 33.38889
```

Thus from the age of 33.39, partial effect of age on education becomes negative.

(c) How would the coefficients and the intercept change if age was instead recorded in months? What would happen to the corresponding t-statistics? Why?

Answer: Lets introduce a new variable $agemth$ where $agemth_i = 12age_i$. The regression then becomes

$$
\begin{aligned}
leduc_i &= \gamma_0 + \gamma_1 IQ_i + \gamma_2 \, agemth_i + \gamma_3 \, agemth_i^2 + \gamma_4 \, black_i + \gamma_5 \, meduc_i + u_i \\
&= \gamma_0 + \gamma_1 IQ_i + \gamma_2 12 age_i + \gamma_3 12^2 age_i^2 + \gamma_4 \, black_i + \gamma_5 \, meduc_i + u_i \\
&= \gamma_0 + \gamma_1 IQ_i + 12\gamma_2 age_i + 144\gamma_3 age_i^2 + \gamma_4 \, black_i + \gamma_5 \, meduc_i + u_i \\
&= \beta_0 + \beta_1 IQ_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 \, black_i + \beta_5 \, meduc_i + u_i
\end{aligned}
$$

Thus the intercept and the coefficients of other variables that have not been transformed do not change. The coefficients of age and age^2 change proportionally where $\gamma_2 = \beta_2/12$ and $\gamma_3 = \beta_3/144$, since the transformations of age variable do not affect the partial effect of age on natural log of education.

The t-statistic is obtained by adjusting the distance of the estimator from its hypothesized value by its standard error:

$$
t = \frac{\text{estimator - hypothesized value}}{\text{st. err. of estimator}} = \frac{\hat{\beta}_j - \beta_{j,\mathbb{H}_0}}{se(\hat{\beta}_j)}
$$

where $se(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$ and is an estimator of the standard deviation of β_j, σ_{β_j}. That is, $se(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2_{\beta_j}}$. The estimator of the variance of β_j, in turn, is

$$
\hat{Var}(\beta_j) = \hat{\sigma}^2_{\beta_j} = \frac{\sigma^2}{TSS_j(1 - R_j^2)} = \frac{\sigma^2}{\left(\sum\limits_{i=1}^{n} (X_{ij} - \bar{X}_j)^2 \right)(1 - R_j^2)}
$$

Since $Var(X) = \frac{1}{n} \sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2$, this can be rewritten as

$$
Var(\beta_j) = \frac{\sigma^2}{nVar(X_j)(1 - R_j^2)}
$$

Since σ is unknown, we replace it with its estimator $\hat{\sigma}$, which gives us the standard error of $\hat{\beta}_j$.

$$
se(\hat{\beta}_j) = \sqrt{\frac{\hat{\sigma}^2}{TSS_j(1 - R_j^2)}} = \sqrt{\frac{\sigma^2}{nVar(X_j)(1 - R_j^2)}}
$$

In this question, the standard error of the estimator γ_2 is then

$$
\begin{aligned}
se(\hat{\gamma}_2) &= \sqrt{\frac{\hat{\sigma}^2}{\left(\sum\limits_{i=1}^{n} (agemth_i - \overline{agemth})^2 \right)(1 - R^2_{agemth})}} \\
&= \sqrt{\frac{\hat{\sigma}^2}{\left(12^2 \sum\limits_{i=1}^{n} (age_i - \overline{age})^2 \right)(1 - R^2_{age})}} \\
&= \sqrt{\left(\frac{1}{12} \right)^2 \hat{Var}(\hat{\beta}_2)}
\end{aligned}
$$

where $\hat{\sigma}^2$ is the estimator of error variance, and R^2_{agemth} is the coefficient of determination from regressing $agemth_i$ on all other explanatory variables. Accordingly, $R^2_{agemth} = R^2_{age}$.

Plugging this into the t-statistic

$$t_{\hat{\gamma}_2} = \frac{\hat{\gamma}_2}{se(\hat{\gamma}_2)} = \frac{\frac{\hat{\beta}_2}{12}}{\frac{1}{12}se(\hat{\beta}_2)} = t_{\hat{\beta}_2}.$$

Thus, the t-statistic for $\hat{\beta}_2$ would not change which would make sense since statistical significance of the explanatory variable should not be impacted by measurement units.

We can apply the same argument to γ_3 with $\gamma_3 = \beta_3/144$.

(d) How would you test that the true coefficient on mother's education is 1/3 the negative of the true coefficient on black?

Answer: The question is asking us to test the following null and alternative hypothesis:

$$\mathbb{H}_0 : \beta_5 = -\frac{1}{3}\beta_4 \quad \text{or} \quad \mathbb{H}_0 : 3\beta_5 + \beta_4 = 0$$

$$\mathbb{H}_1 : \beta_5 \neq -\frac{1}{3}\beta_4 \quad \text{or} \quad \mathbb{H}_1 : 3\beta_5 + \beta_4 \neq 0$$

What we want is to be able to reparameterize the regression so that our null ends up with an estimator equal to 0. For that reparameterization, we can add and subtract $3\beta_5 black_i$ since our null is $3\beta_5 + \beta_4 = 0$ and β_5 is the parameter of $meduc_i$ in the original model specification while β_4 is of $black_i$:

$$leduc_i = \beta_0 + \beta_1 IQ_i + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 black_i + +3\beta_5 black_i + \beta_5 meduc_i - 3\beta_5 black_i + u_i$$

$$= \beta_0 + \beta_1 IQ_i + \beta_2 age_i + \beta_3 age_i^2 + (\beta_4 + 3\beta_5) black_i + \beta_5 (meduc_i - 3black_i) + u_i$$

We can lump $(meduc_i - 3black_i)$ into a new variable for simplicity, say, meb_i. Then we can run regression on the following model:

$$leduc_i = \phi_0 + \phi_1 IQ_i + \phi_2 age_i + \phi_3 age_i^2 + \phi_4 black_i + \phi_5 meb_i + u_i$$

where $meb_i = (meduc_i - 3black_i)$ and $\phi_4 = (\beta_4 + 3\beta_5)$. In this new specification, we would test

$$\mathbb{H}_0 : \phi_4 = 0 \quad \text{which is the same as} \quad \beta_5 = -\frac{1}{3}\beta_4 \quad \text{in the original regression.}$$

$$\mathbb{H}_1 : \phi_4 \neq 0$$

QUESTION 2

Consider the regression model $Y_i = \beta_0 + \beta_1 X_i + U_i$ for an i.i.d. sample with $N = 1,000$ observations. Suppose $U \sim i.i.d.(0, \sigma^2)$ and the X_i are i.i.d. for $i = 1, 2, \ldots, 1000$, and that X_i is independent of U_i. Let $\hat{\beta}_1$ denote the OLS estimator of β_1 and consider another estimator of β_1, $\tilde{\beta}_1$, constructed in the following way:

$$\tilde{\beta}_1 = \frac{Y_3 + Y_1 - 2Y_2}{X_3 + X_1 - 2X_2}.$$

You can assume that X_i are continuously distributed and that $(X_3 + X_1 - 2X_2)^{-1}$ has finite expectation.

(a) Is $\tilde{\beta}_1$ an unbiased estimator of β_1? Why?

(Hint: $\mathbb{E}(\frac{A}{B}) = \mathbb{E}(A \times \frac{1}{B})$ and also if A, B are independent, what can we say about the random variables A and $\frac{1}{B}$?)

Answer: This question is about applying the law of iterated expectations to calculate the expectation of coefficient estimator $\tilde{\beta}_1$ when the regressors are not fixed numbers, but random variables instead.

$\hookrightarrow$ **Important**: Notice that we are assuming that the regressors X_i are *not fixed* but that they are *random variables*. That is, X_i are assumed to be random and are drawn from some distribution. This means, X_i have an expectation and a variance. Importantly, this also means $\mathbb{E}(X_i) \neq X_i$ but $\mathbb{E}(X_i|X_i = x) = x$, or alternatively $\mathbb{E}(X_i|X_i) = X_i$. In other words, conditional on X_i taking a realization x, the expectation of X_i is x.

Unbiasedness requires that the expectation of the estimator is the population parameter. Since X_1, X_2, X_3 are not fixed numbers but random variables, we can't take the unconditional expectation of $\tilde{\beta}_1$. Instead, we need to

- calculate the conditional expectation $\mathbb{E}(\tilde{\beta}_1|X_1, X_2, X_3)$ and

- then utilize the law of iterated expectations.

In order to calculate the first step, first plug in the expression for Y_i in the question into the expression for $\tilde{\beta}_1$ so that we can express it in terms of X_1, X_2, X_3 and U_1, U_2, U_3 :

$$
\begin{aligned}
\tilde{\beta}_1 &= \frac{Y_3 + Y_1 - 2Y_2}{X_3 + X_1 - 2X_2} \\
&= \frac{\beta_0 + \beta_1 X_3 + U_3 + \beta_0 + \beta_1 X_1 + U_1 - 2(\beta_0 + \beta_1 X_2 + U_2)}{X_3 + X_1 - 2X_2} \\
&= \frac{\beta_1(X_3 + X_1 - 2X_2) + (U_3 + U_1 - 2U_2)}{X_3 + X_1 - 2X_2} \\
&= \beta_1 + \frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}.
\end{aligned}
$$

Now we can calculate the conditional expectation:

$$\mathbb{E}(\tilde{\beta}_1|X_1, X_2, X_3) = \mathbb{E}\left(\beta_1 + \frac{U_3 - U_1 - 2U_2}{X_3 + X_1 - 2X_2}|X_1, X_2, X_3\right)$$

$$= \beta_1 + \left(\frac{1}{X_3 + X_1 - 2X_2}\right)\mathbb{E}(U_3 + U_1 - 2U_2|X_1, X_2, X_3)$$

$$= \beta_1 + \left(\frac{1}{X_3 + X_1 - 2X_2}\right)(\mathbb{E}(U_3|X_1, X_2, X_3) + \mathbb{E}(U_1|X_1, X_2, X_3) - 2\mathbb{E}(U_2|X_1, X_2, X_3))$$

$$= \beta_1.$$

The last equality holds because if we recall from the conditional mean expectation assumption, $\mathbb{E}(u_i|X_1, X_2, X_3) = 0$ for $i = 1, 2, 3$.

For the second step, we utilize the law of iterated expectations whereby

$$\mathbb{E}\left(\mathbb{E}(\tilde{\beta}_1|X_1, X_2, X_3)\right) = \mathbb{E}(\tilde{\beta}_1) = \beta_1.$$

Since the expectation of the estimator $\tilde{\beta}_1$ is the parameter β_1, this estimator is unbiased.

(b) Can $\tilde{\beta}_1$ be a better estimator? Why?

Answer: The word *better* means lower variance. To answer this question we will again rely on the law of iterated expectations to first obtain a conditional variance, then unconditional variance. We can then compare that expression of the OLS variance by Gauss-Markov and see if it is better.

Lets start with finding the variance:

$$Var(\tilde{\beta}_1) = Var\left(\beta_1 + \frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)$$

$$= Var\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right) \quad \text{because } \beta_1 \text{ is a constant parameter}$$

$$= \mathbb{E}\left[\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)^2\right] - \left[\mathbb{E}\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)\right]^2 \quad \text{since } Var(A) = \mathbb{E}(A^2) - (\mathbb{E}(A))^2$$

$$= \mathbb{E}\left[\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)^2\right] - \left[\left(\frac{1}{X_3 + X_1 - 2X_2}\right)(\mathbb{E}(U_3) + \mathbb{E}(U_1) - 2\mathbb{E}(U_2))\right]^2$$

$$= \mathbb{E}\left[\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)^2\right] - 0 \quad \text{since } \mathbb{E}(u_i) = 0, \ i = 1, 2, 3$$

$$= \mathbb{E}\left(\frac{U_3^2 + U_1^2 + 4U_2^2 + 2U_3U_1 - 4U_3U_2 - 4U_1U_2}{(X_3 + X_1 - 2X_2)^2}\right).$$

We can now take the conditional expectation of this:

$$\mathbb{E}\left[Var(\tilde{\beta}_1|X_1, X_2, X_3)\right] = \mathbb{E}\left[\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)^2 \bigg| X_1, X_2, X_3\right]$$

$$= \mathbb{E}\left(\frac{U_3^2 + U_1^2 + 4U_2^2 + 2U_3U_1 - 4U_3U_2 - 4U_1U_2}{(X_3 + X_1 - 2X_2)^2} \bigg| X_1, X_2, X_3\right)$$

$$= \mathbb{E}\left(\frac{1}{(X_3 + X_1 - 2X_2)^2} \bigg| X_1, X_2, X_3\right)$$

$$\times \bigg[\mathbb{E}(U_3^2|X_1, X_2, X_3) + \mathbb{E}(U_1^2|X_1, X_2, X_3)$$

$$+ \mathbb{E}(4U_2^2|X_1, X_2, X_3) + \mathbb{E}(2U_3U_1|X_1, X_2, X_3)$$

$$- \mathbb{E}(4U_3U_2|X_1, X_2, X_3) - \mathbb{E}(4U_1U_2|X_1, X_2, X_3)\bigg]$$

$$= \mathbb{E}\left(\frac{1}{(X_3 + X_1 - 2X_2)^2} \bigg| X_1, X_2, X_3\right)\left(\sigma^2 + \sigma^2 + 4\sigma^2 + 0 - 0 - 0\right)$$

$$\text{since } Var(u_i|\vec{X}) = \sigma^2 \text{ , } i = 1, \ldots, n$$

$$\text{and } Cov(u_i, u_j) = \mathbb{E}(u_iu_j) - \mathbb{E}(u_i)\mathbb{E}(u_j) = \mathbb{E}(u_iu_j) = 0$$

$$= \frac{6\sigma^2}{(X_3 + X_1 - 2X_2)^2} \quad \text{since } \mathbb{E}\left(\frac{1}{(X_3 + X_1 - 2X_2)^2} \bigg| X_1, X_2, X_3\right) = \frac{1}{(X_3 + X_1 - 2X_2)^2}.$$

Now we can utilize the law of iterated expectations to derive the unconditional variance:

$$Var(\tilde{\beta}_1) = \mathbb{E}[\mathbb{E}(\tilde{\beta}_1|X_1, X_2, X_3)]$$

$$= \mathbb{E}\left(\mathbb{E}\left[\left(\frac{U_3 + U_1 - 2U_2}{X_3 + X_1 - 2X_2}\right)^2 \bigg| X_1, X_2, X_3\right]\right)$$

$$= \mathbb{E}\left(\frac{6\sigma^2}{(X_3 + X_1 - 2X_2)^2}\right)$$

$$= 6\sigma^2 \mathbb{E}\left(\frac{1}{(X_3 + X_1 - 2X_2)^2}\right).$$

$\hookrightarrow$ Note that we cannot simplify this any further since $\mathbb{E}\left(\frac{1}{(X_3+X_1-2X_2)^2}\right) \neq \frac{1}{(X_3+X_1-2X_2)^2}$ unlike $\mathbb{E}\left(\frac{1}{(X_3+X_1-2X_2)^2} \big| X_1, X_2, X_3\right) = \frac{1}{(X_3+X_1-2X_2)^2}$. This is because X_i are random variables which means $\mathbb{E}(X_i) \neq X_i$ but $\mathbb{E}(X_i|X_i = x) = x$, or alternatively $\mathbb{E}(X_i|X_i) = X_i$, as discussed in part (a).

We can compare this to the variance of $\hat{\beta}_1$:

$$Var(\hat{\beta}_1) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

which we know has the lowest variance among all unbiased linear estimators due to Gauss-Markov theorem.

(c) Can you state the general result for estimators of the form $\sum_{i=1}^{n} a_i Y_i / \sum_{i=1}^{n} a_i X_i$?

Answer: Consider the generalized estimator

$$\beta_1^* = \frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} a_i X_i}$$

where a_i are weights assigned by the estimator to each observation (Y_i, X_i) for $i = 1, \ldots, n$. As usual, we will need to find the expected value and the variance of β_1^*:

$$\mathbb{E}(\beta_1^*) = \mathbb{E}\left(\frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} a_i X_i}\right)$$

$$= \mathbb{E}\left(\frac{\sum_{i=1}^{n} a_i(\beta_0 + \beta_1 X_i + U_i)}{\sum_{i=1}^{n} a_i X_i}\right)$$

$$= \mathbb{E}\left(\frac{\beta_0 \sum_{i=1}^{n} a_i + \beta_1 \sum_{i=1}^{n} a_i X_i + \sum_{i=1}^{n} a_i U_i}{\sum_{i=1}^{n} a_i X_i}\right)$$

$$= \mathbb{E}\left(\beta_0 \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i X_i} + \beta_1 + \frac{\sum_{i=1}^{n} a_i U_i}{\sum_{i=1}^{n} a_i X_i}\right)$$

$$= \beta_0 \mathbb{E}\left(\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i X_i}\right) + \beta_1 + 0 \qquad \text{since } \mathbb{E}(U_i | X) = 0$$

Therefore, β_1^* is unbiased insofar as either $\beta_0 = 0$ or $\mathbb{E}\left(\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} a_i X_i}\right) = 0$.

Notice that $\tilde{\beta}_1$ is one case of β_1^*.

We can derive the general result for the variance of this estimator as well:

$$Var(\beta_1^*) = Var\left(\frac{\sum_{i=1}^{n} a_i Y_i}{\sum_{i=1}^{n} a_i X_i}\right) = Var\left(\frac{\sum_{i=1}^{n} a_i(\beta_0 + \beta_1 X_i + U_i)}{\sum_{i=1}^{n} a_i X_i}\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n} a_i X_i\right)^2} Var\left(\beta_0 \sum_{i=1}^{n} a_i + \beta_1 \sum_{i=1}^{n} a_i X_i + \sum_{i=1}^{n} a_i U_i\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n} a_i X_i\right)^2} \times Var\left(\sum_{i=1}^{n} a_i U_i\right) \qquad \text{since } X_i \text{ and } U_i \text{ are independent}$$

$$= \frac{\sigma_U^2 \sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i X_i\right)^2}.$$

In part (b) we derived $Var(\tilde{\beta})$ as

$$Var(\tilde{\beta}_1) = 6\sigma^2 \mathbb{E}\left(\frac{1}{(X_3 + X_1 - 2X_2)^2}\right)$$

10

which can also be generalized to:

$$Var(\beta_1^*) = \left(\sigma^2 \sum_{i=1}^{n} a_i^2\right) \mathbb{E}\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right].$$

Consequently, the conditional variance is:

$$Var(\beta_1^*|\vec{X}) = \frac{\sigma^2 \sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i X_i\right)^2}$$

where the conditional variance of $\tilde{\beta}_1$ is

$$Var(\tilde{\beta}_1|X_1, X_2, X_3) = \frac{6\sigma^2}{(X_3 + X_1 - 2X_2)^2}.$$

Why OLS weights are the best?

Why OLS weights are the best?

Recall the OLS estimator can be written as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})X_i} = \frac{\sum_{i=1}^{n} a_i^{OLS} Y_i}{\sum_{i=1}^{n} a_i^{OLS} X_i}$$

where the weights are given by the deviation of the regressor from its mean, $a_i = (X_i - \bar{X})$. The OLS weights are the best in the sense that these are the weights that minimize the variance of the general linear unbiased estimator.

We can derive the OLS estimators by setting up the estimation problem as minimization of the variance of the general linear unbiased estimator with respect to the weights of individual observations:

$$\min_{a_1,\dots,a_n} Var(\beta_1^*|\vec{X}) = \min_{a_1,\dots,a_n} \left(\frac{\sigma^2 \sum_{i=1}^{n} a_i^2}{\left(\sum_{i=1}^{n} a_i X_i\right)^2}\right) \quad \text{s.t.} \ \sum_{i=1}^{n} a_i = 0.$$

$\hookrightarrow$ *Note:* From part(a) we know that all the weights need to add up to 0 for unbiasedness.

SUPPLEMENTARY QUESTIONS

QUESTION 1

Using the information given below:

(i) Compute an estimated population correlation coefficient ρ, and the Ordinary Least Squares (OLS) estimates of α and β in the regression of c_t = real consumption expenditure, on y_t = real disposable income, over the period 1948-2003 (dates are inclusive).

$$c_t = \alpha + \beta y_t + \varepsilon_t, \quad t = 1, 2, \ldots \tag{1}$$

where

$$\sum_{t=1}^{n} y_t = 20680 \qquad\qquad \sum_{t=1}^{n} c_t = 19223$$

$$\sum_{t=1}^{n} (y_t - \bar{y})^2 = 1491806 \qquad \sum_{t=1}^{n} (c_t - \bar{c})^2 = 1154516$$

$$\sum_{t=1}^{n} (c_t - \bar{c})(y_t - \bar{y}) = 1309953 \quad \sum_{t=1}^{n} y_t^2 = 9128478$$

$$\sum_{t=1}^{n} c_t^2 = 7752851$$

and $\bar{y}$ and $\bar{c}$ are the sample means of y_t and y_t, respectively.

Answer: First notice that $n = 56$ since the dates are inclusive. Now we need to obtain the OLS estimates and the correlation coefficient:

$$\hat{\beta} = \frac{\sum_{t=1}^{n} (y_t - \bar{y})(c_t - \bar{c})}{\sum_{t=1}^{n} (y - \bar{y})^2} = \frac{1309953}{1491806} = 0.8780988.$$

$$\hat{\alpha} = \bar{c} - \hat{\beta}\bar{y} = \frac{\sum_{t=1}^{n} c_t}{n} - \hat{\beta}\frac{\sum_{t=1}^{n} y_t}{n} = \frac{19223}{56} - 0.8780988\frac{20680}{56} = 18.99853$$

$$\hat{\rho} = \frac{\sum_{t=1}^{n} (c_t - \bar{c})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n} (y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n} (c_t - \bar{c})^2}} = \frac{1309953}{\sqrt{1491806}\sqrt{1154516}} = 0.9981586$$

12

(ii) Show that $\sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n}(c_t - \bar{c})^2 - \hat{\beta}\left(\sum_{t=1}^{n}(c_t - \bar{c})(y_t - \bar{y})\right).$

Answer: Recall that

$$RSS = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n}(c_t - \hat{c}_t)^2$$

$$= TSS - ESS = \sum_{t=1}^{n}(c_t - \bar{c})^2 - \sum_{t=1}^{n}(\hat{c}_t - \bar{c})^2$$

$$= \sum_{t=1}^{n}(c_t - \bar{c})^2 - \sum_{t=1}^{n}\left(\hat{\alpha} + \hat{\beta}y_t - (\hat{\alpha} + \hat{\beta}\bar{y})\right)^2$$

$$= \sum_{t=1}^{n}(c_t - \bar{c})^2 - \sum_{t=1}^{n}\left(\hat{\beta}(y_t - \bar{y})\right)^2$$

$$= \sum_{t=1}^{n}(c_t - \bar{c})^2 - \hat{\beta}^2\sum_{t=1}^{n}(y_t - \bar{y})^2$$

$$= \sum_{t=1}^{n}(c_t - \bar{c})^2 - \hat{\beta}\frac{\sum_{t=1}^{n}(y_t - \bar{y})(c_t - \bar{c})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}\sum_{t=1}^{n}(y_t - \bar{y})^2 \quad \text{plugged-in definiton of } \hat{\beta}$$

$$= \sum_{t=1}^{n}(c_t - \bar{c})^2 - \hat{\beta}\left(\sum_{t=1}^{n}(y_t - \bar{y})(c_t - \bar{c})\right)$$

as desired. We can plug-in the values to obtain:

$$\sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n}(c_t - \bar{c})^2 - \hat{\beta}\left(\sum_{t=1}^{n}(c_t - \bar{c})(y_t - \bar{y})\right)$$

$$= 1154516 - 0.8780988 \times 1309953 = 4247.843$$

(iii) Test separately the hypotheses that α, β, and ρ are zero, at the 5 percent significance level. (Note: the t-statistic for your test of β and ρ should be exactly the same in this case - check that it is)

Answer: Since the hypothesis is testing if they are individually equal to zero or not, we will use two-tailed t-test. Recall that

$$\frac{\bar{X} - \mu_X}{\sqrt{\dfrac{s_X^2}{n_X}}} \sim N(0,1).$$

To apply this to each of α, β, and ρ, we need to first find their standard errors. For that, we need to find their variances. We derived these in class notes under the discussion of "Precision of OLS Estimators: Standard Errors".

$$se(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^2}{\sum_{t=1}^{n}(y-\bar{y})^2}} = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^2}{1491806}} \quad , \quad se(\hat{\alpha}) = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^2 \sum_{t=1}^{n} y_t^2}{n \sum_{t=1}^{n}(y-\bar{y})^2}} = \sqrt{\frac{\hat{\sigma}_{\varepsilon}^2 \times 9128478}{56 \times 1491806}}$$

so we will need to find the variance of the residuals next:

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{4247.843}{54} = 78.66376$$

Plugging this back in, we have

$$se(\hat{\beta}) = \sqrt{\frac{78.66376}{1491806}} = 0.007261581 \quad , \quad se(\hat{\alpha}) = \sqrt{\frac{78.66376 \times 9128478}{56 \times 1491806}} = 2.931814$$

Now we can calculate the t-statistic with the values we calculated in part (i):

$$t_{\hat{\alpha}} = \frac{\hat{\alpha} - \mu_{\alpha}}{se(\hat{\alpha})} = \frac{18.99853 - 0}{2.931814} = 6.480128$$

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \mu_{\beta}}{se(\hat{\beta})} = \frac{0.8780988 - 0}{0.007261581} = 120.9249$$

for both we can then reject the null hypothesis that they are individually equal to zero.

To understand ρ, recall that $\varepsilon_t = c_t - \hat{c}_t$ where $\hat{c}_t = \hat{\alpha} + \hat{\beta} y_t$. That is, the randomness in c_t comes from two sources: y_t and ε_t. That is, the variance of the dependent variable is:

$$Var(c_t) = \beta^2 Var(y_t) + Var(\varepsilon_t),$$

This means the total variation in c_t can be divided into two parts. The first part, $\beta^2 Var(y_t)$ is the variation due to y_t, while the second part, $Var(\varepsilon_t)$, is the variance of error, which is the variance remaining after knowing y_t.

What we are then looking for is the proportion of variance of c_t that is explained by variation in y_t, which is the definition of covariance between c_t and y_t. That is, the closer both covary, the more c_t can be approximated by a linear function of y_t: i.e. $c_t \approx \hat{c}_t = \alpha + \beta y_t$:

$$\frac{\beta^2 Var(y_t)}{Var(c_t)} = \frac{\left(\frac{Cov(c_t, y_t)}{Var(y_t)}\right)^2 Var(y_t)}{Var(c_t)} = \frac{\left(Cov(c_t, y_t)\right)^2}{Var(c_t)} = \rho^2.$$

Since by definition, $\rho = \frac{Cov(c_t, y_t)}{\sqrt{Var(c_t) Var(y_t)}}$.

Thus, ρ^2 indicates the strength of our regression model in estimating c_t from Y_t. In practice we do not have ρ but we have the observed pairs $(c_1, y_1), (c_2, y_2), \ldots, (c_n, y_n)$ and we estimate ρ from the observed data, which is the definition of R-squared, or coefficient of determination.

This means, β can also be expressed as:

$$\beta = \sqrt{\frac{\rho^2 Var(c_t)}{Var(y_t)}} = \rho \frac{\sigma_{c_t}}{\sigma_{y_t}}$$

Therefore we can rewrite our model as:

$$c_t = \alpha + \beta y_t + \varepsilon_t$$
$$= \left(\bar{c} - \rho \frac{\sigma_{c_t}}{\sigma_{y_t}} \bar{y} \right) + \left(\rho \frac{\sigma_{c_t}}{\sigma_{y_t}} \right) y_t$$

The standard error of ρ was first derived by Pearson using maximum likelihood.

$$se(\hat{\rho}) = \sqrt{\frac{1 - \hat{\rho}^2}{n - 2}} = \sqrt{\frac{1 - 0.9981586^2}{54}} = 0.008254527$$

The t-statistic is therefore:

$$t_{\hat{\rho}} = \frac{\hat{\rho} - \rho^0}{se(\hat{\rho})} = \frac{0.9981586 - 0}{0.008254527} = 120.9226$$

as desired. Since it is way more than 2 we can reject the null hypothesis.

it is way more than 2 we can reject the null hypothesis.

\end{description}

(iv) Test the hypothesis that β is zero again, now using the following expression for the F-statistic:

$$F = \frac{\dfrac{(RSS_R - RSS_U)}{df}}{\dfrac{RSS_U}{df}} \tag{2}$$

where RSS_R and RSS_U are the residual sum of squares of the restricted and unrestricted equation respectively and df are degrees of freedom.

(**Note:** when testing the same hypothesis, the F-statistic is equal to the t-statistic squared - use this to check your answers to (iii) and (iv)).

Answer: Notice that we are only restricting β so the numerator degrees of freedom is $q = df_R - df_U = 1$ and the denominator degrees of freedom is $n - k - 1 = 56 - 1 - 1 = 54$.

Now we just need to find the RSS for unrestricted and restricted. For unrestricted, this is our usual:

$$RSS_U = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (c_t - \bar{c})^2$$

For the restricted one, the model is $c_t = \alpha + \varepsilon_t$ thus it only depends on the error. Accordingly,

$$RSS_R = \sum_{t=1}^{n} e_t^2$$

15

Therefore, the F-test is:

$$F = \frac{\dfrac{(RSS_R - RSS_U)}{df}}{\dfrac{RSS_U}{df}} = \frac{\dfrac{\left(\sum\limits_{t=1}^{n}(c_t - \bar{c})^2 - \sum\limits_{t=1}^{n} e_t^2\right)}{1}}{\dfrac{\sum\limits_{t=1}^{n} e_t^2}{54}} = \frac{\dfrac{1154516 - 4247.843}{1}}{\dfrac{4247.843}{54}} = 14622.59$$

To verify, if we take the squareroot of this:

$$\sqrt{14622.59} = 120.9239$$

as desired.

— —

QUESTION 2

In the linear regression model $y_t = a + bt + e_t$ where $t = 1, \ldots, T$ (i.e. t is a linear trend, and so non-stochastic) and e_t is an independently distributed random variable with zero mean and constant variance σ^2, show that the estimator b^* defined by $b^* = \frac{(y_T - y_1)}{T-1}$ is or is not, an unbiased estimator of b. Use this result to comment on the claim that b^* has a lower variance that the Ordinary Least Squares estimator. Explain your answer. Calculate the variance of b^*. (Optional question: for what values of T would the estimators have the same variance?)

Answer: Lets start with obtaining an expression for the estimator b^*:

$$b^* = \frac{(Y_T - y_1)}{T - 1} = \frac{(a + bT + e_T) - (a + b1 + e_1)}{T - 1} = \frac{b(T - 1) + e_T - e_1}{T - 1}$$

So, b^* can only be unbiased if $e_T = e_1 = 0$. Then, $\mathbb{E}(b^*) = b$.

Next we will look at the variance of b^*:

$$Var(b^*) = Var\left(\frac{y_T - y_1}{T - 1}\right) = \left(\frac{1}{T - 1}\right)^2 Var(y_T - y_1) = \frac{2\sigma_e^2}{(T - 1)^2}$$

Lets compare this to OLS estimator's variance:

$$Var(\hat{\beta}) = \frac{\sigma_e^2}{\sum\limits_{t=1}^{n}(t - \bar{t})^2}$$

where

16

$$\sum_{t=1}^{T}(t-\bar{t})^2 = \sum_{t=1}^{T}(t-\bar{t})(t-\bar{t})$$

$$= \sum_{t=1}^{T}(t-\bar{t})t$$

$$= \sum_{t=1}^{T}t^2 - \bar{t}\sum_{t=1}^{T}t \quad \text{and since } \sum t^2 = \frac{T(T+1)(2T+1)}{6} \text{ and } \sum t = \frac{T(T+1)}{2}$$

$$= \frac{T(T+1)(2T+1)}{6} - \left(\frac{T(T+1)}{2T}\right)\left(\frac{T(T+1)}{2}\right)$$

$$= \frac{2T(T+1)(2T+1) - 3T(T+1)^2}{12}$$

$$= \frac{T(T+1)(T-1)}{12}$$

So, plugging this back in to the variance expression:

$$Var(\hat{\beta}) = \frac{12\sigma_e^2}{T(T+1)(T-1)}$$

Thus, $Var(\hat{\beta}) \leq Var(b^*)$ for $T > 1$. The two can be equal if $\frac{2\sigma_e^2}{(T-1)^2} = \frac{12\sigma_e^2}{T(T+1)(T-1)}$ which can happen only if

$$\frac{1}{T-1} = \frac{6}{T(T+1)}$$

$$T(T+1) = 6(T-1)$$

$$T^2 - 5T + 6 = 0$$

$$(T-2)(T-3) = 0$$

$$T = 2 \text{ or } T = 3.$$

_ _

QUESTION 3

Briefly discuss the 'ceteris paribus' and 'partialling out' interpretations of the coefficients estimated using multiple regression analysis.

Answer: Consider the case of an *estimated equation* with $k = 2$ independent variables: $\hat{Y} = \hat{\beta}_0 + \hat{(\beta)}_1 X_1 + \hat{\beta}_2 X_2$.

Ceteris paribus: The estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have partial effect interpretations. From our estimated equation we have $\Delta \hat{Y} = \hat{\beta}_1 \Delta X_1 + \hat{\beta}_2 \Delta X_2$ which tells us that we can obtain the predicted change in Y given the changes in X_1 and X_2. Notice we do not have the intercept here, since intercept is the predicted value of Y when both X_1 and X_2 are zero. The coefficient on X_1 measures the change in $\hat{Y}$ due to a one unit change in X_1, holding all other independent variables fixed: $\Delta \hat{Y} = \hat{\beta}_1 \Delta X_1$ holding X_2 fixed - i.e. $\Delta X_2 = 0$. The case with $k > 2$ is similar. This way, we have *controlled for* the variables $X_2, \ldots, X_k$ when estimating the effect of X_1 on Y. The other coefficients have a similar interpretation.

Another way of thinking about this is to look back how we derived $\hat{\beta}_1$ before. We took the partial derivative of of Y with respect to X_1, $\hat{\beta}_1 = \frac{\partial Y}{\partial X_1}$. This tells us the rate of change of Y with respect to X_1 holding X_2 constant.

Partialling Out: This is another representation of the partial effect interpretation of $\hat{\beta}_1$. For this approach, we partial out out the effect of X_2 by first regressing X_1 on X_2 using the model $X_1 = \alpha_0 + \alpha_1 X_2 + r_1$. Here, the residual r_1 represents all the influences on X_1 not related to X_2. If we then regress Y on this residual using the model $Y = \gamma_0 + \gamma_1 r_1 + v$, then Frisch, Waugh, Lovell Theorem tells us that $\hat{\gamma}_1 = \hat{\beta}_1$. That is, $\hat{\beta}_1 = \frac{Cov(Y,v)}{Var(v)}$. Thus, this shows that $\hat{\beta}_1$ is estimating the effect of X_1 on Y after the effect of X_2 has been partialled out.

– –

QUESTION 4

Consider the model
$$y_t = \alpha + \beta_1 X_{t1} + \beta_2 X_{t2} + \varepsilon_t \quad , \quad t = 1, 2, 3, \ldots, n$$

where: y_t = logarithm of output, x_{t1} = logarithm of labor input, and x_{t2} = logarithm of capital input.

The latter two are the exogenous variabes and the ε_t are independently $N(0, \sigma^2)$. The estimated equation (using the OLS method) based on the data over the period 1929 - 1967 (inclusive) is

$$\hat{y}_t = 3.38466 + 1.4100 x_{t1} + 0.4162 x_{t2} \quad , \quad \text{where } R^2 = 0.5910.$$

The variance-covariance matrix of the least squares estimates of β_1 and β_2 is estimated to be:

$$\begin{bmatrix} 0.7820 & -0.4235 \\ -0.4235 & 0.2549 \end{bmatrix}$$

i.e. $Var(\hat{\beta}_1) = 0.7820$, $Var(\hat{\beta}_2) = 0.2549$, $Cov(\hat{\beta}_1, \hat{\beta}_2) = -0.4235$.

(a) Test the following hypothesis separately, each at 0.05 level:

$$\mathbb{H}_1 : \beta_1 = 0 \quad \textbf{and} \quad \mathbb{H}_2 : \beta_2 = 0.$$

Answer: Given that we are testing these separately, and the alternative hypotheses are that they are not equal to zero, we need to use two-tailed t-test for each.

Since the number of coefficient estimators are 3, we have $n - 3$ degrees of freedom and $n = 39$, so the critical values are $t_{0.025,36} = -2.028$ and $t_{0.975,36} = 2.028$.

For β_1 hypothesis, the t-statistic is

$$t_{\beta_1} = \frac{\hat{\beta}_1}{\sqrt{Var(\hat{\beta}_1)}} = \frac{1.41}{\sqrt{0.782}} = 1.59447$$

Similarly, for β_2 hypothesis, the t-statistic is

$$t_{\beta_2} = \frac{\hat{\beta}_2}{\sqrt{Var(\hat{\beta}_2)}} = \frac{0.4162}{\sqrt{0.2549}} = 0.82436$$

Since neither of them exceed the absolute value of the critical value we cannot reject the null hypothesis.

(b) Given that R^2 is defined as ESS/TSS, and that $TSS = RSS + ESS$ where ESS is the explained sum of squares, TSS is the total sum of squares, and RSS is the residual sum of squares, show that equation 2 in Question 1(iv) above, can be written as:

$$F = \frac{\frac{(R_U^2 - R_R^2)}{df}}{\frac{1 - R_U^2}{df}}. \tag{3}$$

Under what conditions can equation 2 not be rearranged this way?

Answer: F is defined in equation 2 as:

$$F = \frac{\frac{(RSS_R - RSS_U)}{df}}{\frac{RSS_U}{df}}.$$

Since $R^2 = 1 - \frac{RSS}{TSS}$, we have $RSS = (1 - R^2)TSS$. Plugging this into the definition of F, we get:

$$F = \frac{\dfrac{\left((1 - R_R^2)TSS_R - (1 - R_U^2)TSS_U\right)}{df}}{\dfrac{(1 - R_U^2)TSS_U}{df}}$$

$$= \frac{\dfrac{\dfrac{1}{TSS}\left((1 - R_R^2)TSS - (1 - R_U^2)TSS\right)}{df}}{\dfrac{1}{TSS}\dfrac{(1 - R_U^2)TSS}{df}} \qquad \text{since } TSS_R = TSS_U$$

$$= \frac{\dfrac{1 - R_R^2 - 1 + R_U^2}{df}}{\dfrac{1 - R_U^2}{df}}$$

$$= \frac{\dfrac{(R_U^2 - R_R^2)}{df}}{\dfrac{1 - R_U^2}{df}}$$

as desired. Alternatively, we can use $R^2 = \frac{ESS}{TSS}$ or $ESS = R^2 \times TSS$ and $RSS = TSS - ESS$ identities:

$$F = \frac{\dfrac{(RSS_R - RSS_U)}{df}}{\dfrac{RSS_U}{df}} = \frac{\dfrac{(TSS_R - ESS_R) - (TSS_U - ESS_U)}{df}}{\dfrac{TSS_U - ESS_U}{df}}$$

$$= \frac{\dfrac{(TSS_R - R_R^2 \times TSS_R) - (TSS_U - R_U^2 \times TSS_U)}{df}}{\dfrac{TSS_U - R_U^2 \times TSS_U}{df}}$$

$$= \frac{\dfrac{\left(TSS_R(1 - R_R^2)\right) - \left(TSS_U(1 - R_U^2)\right)}{df}}{\dfrac{TSS_U(1 - R_U^2)}{df}}$$

$$= \frac{\dfrac{TSS(R_U^2 - R_R^2)}{df}}{\dfrac{TSS_U(1 - R_U^2)}{df}} \qquad \text{since } TSS_R = TSS_U$$

$$= \frac{\dfrac{(R_U^2 - R_R^2)}{df}}{\dfrac{1 - R_U^2}{df}}$$

as desired.

This is called the R-**squared form of the F statistic**. Since R^2 is reported with almost all regressions while RSS is not, it is easier to use R^2s from the unrestricted and restricted models to test for exclusion of some variables. Notice that in the numerator the *unrestricted* R^2 comes first unlike the numerator in equation 2. Since $R_U^2 > R_R^2$, the F statistic will always be positive.

Even though R^2-form of the F statistic is convenient for testing exclusion restrictions, it can't be applied for testing all linear restrictions. Suppose we want to test whether labor is a rational valuation. If this is the case, then a 1% change in X_1 should be associated with a 1% change in the output. That is, $\beta_1 = 1$. In addition, capital should not help to explain the output once labor has been controlled for. The hypothesis is then:

$$\mathbb{H}_0 : \beta_1 = 1, \beta_2 = 0.$$

We therefore have one exclusion restriction and $\beta_1 = 1$ is not. To test the hypothesis using F statistic, we first estimate the unrestricted model and then impose the restrictions in the hypothesis to obtain the restricted model.

Here, however, the second step can be a bit tricky. The restricted model is $y_t = \alpha + 1x_{t1} + 0\beta_2 x_{t2} + \varepsilon$. To impose restriction that the coefficient on x_{t1} is unity, we need to estimate the following model:

$$y_t - x_{t1} = \alpha + \varepsilon$$

which is a model with just an intercept but with a different dependent variable than the unrestricted model. Because of this difference in the dependent variable between unrestricted and restricted models, we cannot in this case use R^2 form of the F statistic. As a result, the TSS_U and TSS_R would no longer be equal and thus the R-squared form of F statistic is not equal to its RSS form.

In general, the RSS form of F statistic should be used if different dependent variable is needed in running the restricted regression.

(c) If an eqaution is restricted such that all slope coefficiencts are zero, what is the coefficient of determination in R^2 of that equation?

Answer: Recall the definition of R^2:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{t=1}^{n} \hat{\varepsilon}_t^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}.$$

If all the slope coefficients are zero, then the restricted model is just the intercept, so both the numerator and denominator becomes the variance of dependent variable. Accordingly, $R^2 = 1 - 1 = 0$.

We can also arrive to the same conclusion by looking at the restricted model:

$$\hat{y}_t = \alpha + \varepsilon_t$$

where $\alpha = \bar{y}$. We can now use this within the expression for R^2:

$$R^2 = \frac{ESS}{TSS}$$

$$= \frac{\sum_{t=1}^{n}(\hat{y} - \bar{y})^2}{TSS}$$

$$= \frac{\sum_{t=1}^{n}(\bar{y} - \bar{y})^2}{TSS}$$

$$= \frac{0}{TSS}$$

$$= 0.$$

Intuitively, having an intercept-only model means we have a horizontal line equal to α. Since R^2 quantifies how well x is associated with y, then this means knowing x does not help in understanding y. Therefore, the best prediction of y is the mean of all y_t irrespective of x. Since x has no useful information in understanding y, the R^2 would be zero.

(d) Using equation 3 and your answer to part (c) above, test the joint hypothesis at the .05 **level**

$$\mathbb{H}_3 : \beta_1 = \beta_2 = 0$$

Answer: Since it is a joint hypothesis we will use the F test by plugging in the values. We know from the question $R^2_U = 0.5910$ and from part (c) $R^2_R = 0$. We then just need to determine the degrees of freedom. The numerator df is the difference between the number of explanatory variables in the unrestricted and the restricted models which, in this case, is $2 - 0 = 2$.

The denominator df is the difference between the sample size and the degrees of freedom used up, which, in this case, is $38 - 2 = 36$. Therefore,

$$F = \frac{\frac{(R^2_U - R^2_R)}{df}}{\frac{1 - R^2_U}{df}} = \frac{\frac{(0.5910 - 0)}{2}}{\frac{1 - 0.5910}{36}} = 26.00978$$

The critical value is: $F_{(0.05,2,36)} = 3.25945$. Since F-statistic is higher than the critical value, we can reject the null hypothesis at 0.05.

(e) Show that $Var(\hat{\beta}_1 + \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_2)$**, and use this result to test the hypothesis that** $\beta_1 + \beta_2 = 1$ **using the** t**-distribution (your test statistic here should be roughly 1.896).**

Answer: Since $Var(X) = \mathbb{E}\left(X - \mathbb{E}(X))^2\right)$ we can derive variance of the sum of two βs as follows:

$$
\begin{aligned}
Var(\hat{\beta}_1 + \hat{\beta}_2) &= \mathbb{E}\left(\left((\hat{\beta}_1 + \hat{\beta}_2) - \mathbb{E}(\hat{\beta}_1 + \hat{\beta}_2)\right)^2 \right) \\
&= \mathbb{E}\left(\left(\left[\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1)\right] + \left[\hat{\beta}_2 - \mathbb{E}(\hat{\beta}_2)\right]\right)^2 \right) \\
&= \mathbb{E}\left(\left[\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1)\right]^2 + \left[\hat{\beta}_2 - \mathbb{E}(\hat{\beta}_2)\right]^2 + 2\left[\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1)\right]\left[\hat{\beta}_2 - \mathbb{E}(\hat{\beta}_2)\right] \right) \\
&= \mathbb{E}\left(\left[\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1)\right]^2 \right) + \mathbb{E}\left(\left[\hat{\beta}_2 - \mathbb{E}(\hat{\beta}_2)\right]^2 \right) + 2\mathbb{E}\left(\left[\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1)\right]\left[\hat{\beta}_2 - \mathbb{E}(\hat{\beta}_2)\right] \right) \\
&= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_2).
\end{aligned}
$$

Our hypothesis is:

$$
\begin{aligned}
\mathbb{H}_4 &: \beta_1 + \beta_2 = 1 \\
\mathbb{H}_A &: \beta_1 + \beta_2 \neq 1.
\end{aligned}
$$

which means we will use 2-tailed t-test. The critical values are $t_{0.025,36} = -2.0281$ and $t_{0.975,36} = 2.0281$, and the t-statistic is:

$$
t = \frac{\text{our estimate - hypothesized value}}{\text{standard error}} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{\sqrt{Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_2)}}
$$

$$
= \frac{(1.4100 + 0.4162) - 1}{\sqrt{0.7820 + 0.2549 + 2(-0.4235)}} = 1.895932.
$$

Since the t-statistic $= 1.8956 < 2.0281 =$ critical value, we cannot reject the null hypothesis.

(f) Interpret the above results. In particular, how do you interpret the result in part (e) abouve and how do you explain the differences in your results to parts (a) and (d)?

Answer: In part (a) we tested whether β_1 or β_2 are separately equal to 0 using t-test and could not reject either of the nulls at 0.05 level. In part (d) we tested a third hypothesis of whether both βs are jointly equal to 0 using F test and rejected this null at 0.05 level. Finally, in part (e) we tested if the βs add up to 1, and could not reject the null.

So in part (d) F statistic is significant and in part (a) the t statistics are not. This can happen if the explanatory variables are *highly correlated* and the model is subject to strong or severe multicollinearity.

If this happens, the standard errors of the slope coefficients could all be so large that none of the t-statistics is significant. Yet, if the model is correctly specified and R^2 is high then it is likely the F statistic is significant. In this situation, we'd know that our model has high explanatory power, but we wouldn't be in a position to isolate the contributions made by the explanatory variables individually.

Also, since $\beta_1 + \beta_2 = 1$ cannot be rejected, we can interpret it as a *constant returns to scale*. That is, the proportional increase in the output and the inputs is the same.

– –

QUESTION 5

An economist wishes to estimate the following model from aggregate time series:

$$q = a_0 + a_1 y + a_2 p_1 + a_3 p_2 + a_4 n_u$$

where q is the volume of food consumption, y real disposable income, p_1 an index of the price of food, p_2 an index of all other prices and n population. All variables are in logarithms.

The correlation between p_1 and p_2 is found to be 0.95 and that between y and n is found to be 0.93. The economist assumes that the equation suffers from the problem of multicollinearity and asks colleague for advice.

(A) **suggests dropping all variables with t-statistics less than 2.**

(B) **says that multicollinearity results from too little data variation and suggests pooling the aggregate time series data with a cross-section budget survey on food consumption.**

(C) **suggets that the amount being asked of the data could be reduced by imposing the restrictions $a_2 + a_3 = 0$ and $1 - a_1 - a_4 = 0$ which are suggested by economic theory.**

(D) **says multicollinearity will be reduced by replacing equation** with equation below:

$$Z_i = \beta_0 + \beta_1 Z_2 + \beta_2 Z_3 + \beta_3 Z_4 + u$$

where $Z_i = q - n, Z_2 = y - n, Z_3 = p_1 - p_2, Z_4 = p_1$, because in equation the correlations between the right hand side variables are lower.

(E) **says that adding lagged values of q to the equation will reduce multicollinearity since it is known that it has a significant effect on food consumption.**

(i) State exactly what you understand by the term multicollinearity and suggest ways in which it might be possible to detect whether multicollinearity, understood in this way, can be detected.

Answer: Multicollinearity, or imperfect multicollinearity, is a <u>sample</u> problem and means that two or more of the regressors are highly correlated. This means, there is a linear function of the regressors that is highly correlated with another regressor. To see this, consider the sampling variance of OLS slope estimator:

$$Var(\hat{a}_j) = \frac{\sigma_u^2}{TSS_j(1 - R_j^2)} = \frac{\frac{1}{n}\sum_{i=1}^{n} u_i^2}{TSS_j(1 - R_j^2)}$$

which has three components:

σ_u^2 : the larger the error variance the bigger the variance of OLS slope estimator. It is the one component of the equation above that is unknown. For a given dependent variable, there is only one way to reduce σ_u^2: add more explanatory variables to the equation, which would take out some factors out of the error term. However, it is not always possible to find additional legitimate factors that affect q.

TSS_j : From the equation above we can see that the larger the total sample variation in the jth explanatory variable the smaller the variance of the OLS slope estimator will be. We can increase the TSS_j by increasing the sample size.

R_j^2 : This is obtained from regressing the jth independent variable against all the other independent variables. It shows the proportion of the total variation in the jth independent variable that can be explained by the other independent variables.

For a given σ_u^2 and TSS_j, the extreme case of $R_j^2 = 1$ means perfect multicollinearity and is a violation of our OLS and Gauss-Markov assumptions. Whereas if $R_j^2 = 0$, which can occur if and only if the jth variable is has zero sample correlation with every other independent variable, then the smallest $Var(\hat{a}_j)$ is obtained. This is the best case but rarely encountered.

When $R_j^2 < 1$, but "close" to 1, then we have imperfect multicollinearity, or just multicollinearity However, there is not a precise definition of "close" in this context. For example $R_j^2 = 0.9$ means 90% of the sample variation in the jth independent variable can be explained by other independent variables in the regression model. However, whether this translates into a $Var(\hat{a}_j)$ that is too large to be useful depends on the sizes of σ_u^2 and TSS_j.

As well as a large R_j^2, a small sample size can lead to a small value of TSS_j and therefore can lead to large sampling variances as well. Therefore large R_j^2 and small sample size generate much the same problem in this context: large OLS estimator variance. This in turn means large standard error, which can cause the t-statistic for the coefficient to become not significant.

Overall, it is better to have less correlation between the jth independent variable and the other independent variables for estimating $\hat{a}_j$. However, there is no great way of tackling multicollinearity. In social sciences where data are usually collective passively, there is no good way to reduce variances of unbiased estimators other than to collect more data. We can try to drop other independent variables from the model to try to reduce multicollinearity but dropping a variable that belongs in the population model can lead to bias.

There are some methods to determine the severity of multicollinearity, such as variance inflation factor or, VIF, but their use is limited.

The VIF for slope coefficient is

$$VIF_j = \frac{1}{1 - R_j^2}$$

which allows us to write the variance of the jth OLS estimator as:

$$Var(\hat{a}_j) = \frac{\sigma_u^2}{SST_j} \times VIF_j.$$

This shows that VIF_j is the factor by which $Var(\hat{a}_j)$ is higher. However, looking at the size of VIF_j is of limited use since any cut-off point would be arbitrary.

Another way to detect multicollinearity is to compare t-tests with F-tests and R^2 as we did in Question 4.

(ii) What problems might regression analysis have if multicollinearity is present? Is the economist correct in assuming these problems will be present in this case?

Answer: The economist is not correct. It is the case that multicollinearity is present but it is not necessarily a problem. It depends on the sample. If the regressors are imperfectly multicollinear, then the coefficients on at least one individual regressor will be imprecisely estimated. Given the correlation between p_1 and p_2 is 0.95 for example, it would be difficult to use these data to estimate the a_2 holding constant the index of all other prices. As a result, the OLS estimator of a_2 will have a larger variance than if the regressors p_1 and p_2 were uncorrelated. Assuming the model is correctly specified, presence of multicollinearity implies that it will be difficult to estimate precisely one or more of the partial effects using the data at hand.

(iii) Suppose multicollinearity *is* a problem in this case. Commnet on the appropriateness in each of the suggestions made above.

Answer: (A): This is not an appropriate approach. Even if $t < 2$ the variable we drop may be important. Leaving a relevant variable with non-zero coefficient out of the model results in a biased estimator of the slope coefficient of the remaining variable.

(B): this can work in that increasing sample size is good. As the sample size increases, the variance of the estimator decreases. This is also why we wouldn't want to drop a variable. If we drop a variable, it is true that the variance of the remaining estimator is lower. However, we can reduce the increase in variance caused by including a correlated regressor by increasing the sample size. The bias created by dropping a variable, on the other hand, is independent of the sample size. Note: There are potential problems of combining time series and cross-section data but these will be discussed after covering time series.

(C): Yes, this would reduce the amount required of equation and amount asked of the data. This seem to be in keeping with the economic theory: $\frac{q}{n} = e^{a_0} + \left(\frac{y}{n}\right)^{a_1} \left(\frac{p_1}{p_2}^{a_2}\right)$.

(D): Similar to (C) in principle, but it is not in keeping with the economic theory. So it may not be that interpretable.

(E): This one is a bit messy. It may make multicollinearity worse, but it may alleviate the problem as it would reduce $\hat{\sigma}_u^2$.